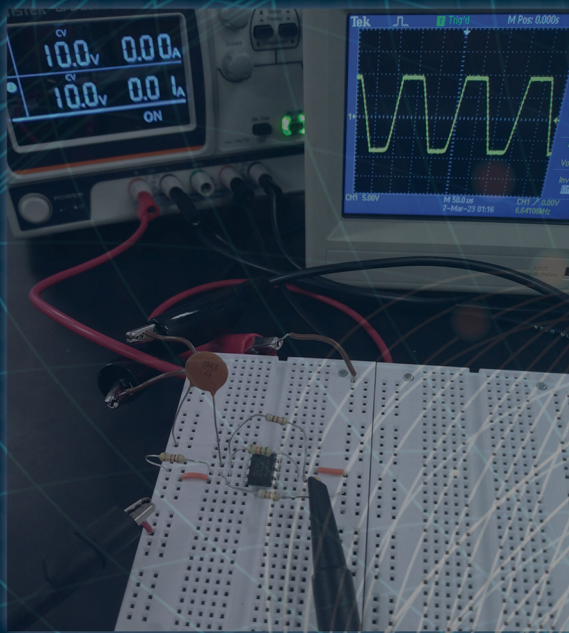


ELECTRONICS for SCIENTISTS



DANIEL F. SANTAVICCA



CRC Press
Taylor & Francis Group

VISIT...

LANZAROTE
Caliente.COM

Electronics for Scientists

Electronics for Scientists provides a practical and concise introduction to electrical circuits, signals, and instrumentation for undergraduate students in the physical sciences. No previous familiarity with electronics is required, and concepts are grounded in the relevant physics. The book aims to give students the electronics background needed to be successful in experimental science.

The book begins with the fundamentals of DC circuits. This is followed by AC circuits and their analysis using the concept of impedance. The transfer function is introduced and used to analyze different types of filter circuits. The conversion between time-domain and frequency-domain signal representations is reviewed. Transmission lines are introduced and used to motivate the different approaches to designing microwave-frequency circuits as compared to lower-frequency circuits. The physics of semiconductors is reviewed and used to understand the behavior of diodes and transistors, and a number of diode and transistor circuits are analyzed. The operational amplifier (op-amp) is introduced and several op-amp circuits are analyzed. Techniques for quantifying noise in electrical measurements are described and common sources of noise are discussed. The last major topic is digital circuits, which include analog-to-digital conversion, logic gates, and digital memory circuits. The book concludes with a brief introduction to quantum computing.

Designed for a one-semester course, this book brings together a range of topics relevant to experimental science that are not commonly found in a single text. Worked examples are provided throughout the book, and each chapter concludes with a set of problems to reinforce the material covered. The subject of electronics is indispensable to a wide array of scientific and technical fields, and this book seeks to provide an approachable point of access to this rich and important subject.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Electronics for Scientists

Daniel F. Santavicca



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

First edition published 2024
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

© 2024 Daniel F. Santavicca

Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access www.copyright.com or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC please contact mpkbookspermissions@tandf.co.uk

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

ISBN: 978-1-032-52812-0 (hbk)
ISBN: 978-1-032-52813-7 (pbk)
ISBN: 978-1-003-40849-9 (ebk)

DOI: [10.1201/9781003408499](https://doi.org/10.1201/9781003408499)

Typeset in Nimbus Roman font
by KnowledgeWorks Global Ltd.

Publisher's note: This book has been prepared from camera-ready copy provided by the authors.

Contents

Preface.....	ix
Author	xi
Chapter 1	Linear DC Circuits 1
1.1	Ohm's law 1
1.2	Introduction to circuit diagrams..... 3
1.3	Power in DC circuits..... 9
1.4	Kirchhoff's rules 10
1.5	Sources and meters 12
1.6	Thevenin and Norton equivalent circuits 16
1.7	Problems 18
Chapter 2	Linear AC Circuits..... 23
2.1	Capacitance and inductance..... 23
2.2	Transient analysis 26
2.2.1	RC circuit..... 26
2.2.2	RL circuit 28
2.2.3	LC circuit 28
2.3	Impedance..... 30
2.4	Power in AC circuits..... 33
2.5	Resonant circuits..... 34
2.5.1	Series RLC circuit..... 34
2.5.2	Parallel RLC circuit 36
2.6	The oscilloscope 38
2.7	Problems 41
Chapter 3	Four-Terminal Circuits and the Transfer Function 45
3.1	The transfer function..... 45
3.2	Simple filter circuits..... 45
3.3	Fourier analysis..... 49
3.3.1	Fourier series..... 50
3.3.2	Fourier transform 51
3.4	Transformers 53
3.5	Problems 55

Chapter 4	Transmission Lines	59
4.1	Lumped-element model	59
4.2	Terminated transmission lines	64
4.3	Special topic: scattering parameters	66
4.4	Special topic: the half-wave dipole antenna	67
4.5	Problems	70
Chapter 5	Semiconductor Devices	71
5.1	Physics of semiconductors.....	71
5.2	PN junction	75
5.3	Diode circuits.....	79
5.4	Frequency mixing	84
5.5	Special topic: the lock-in amplifier.....	86
5.6	Transistors.....	86
5.6.1	NPN bipolar junction transistor	87
5.6.2	Transistor circuits.....	89
5.7	Special topic: integrated circuits.....	95
5.8	Problems	97
Chapter 6	Operational Amplifiers	101
6.1	Ideal op-amps with stable feedback.....	102
6.2	Nonidealities in real op-amps	105
6.3	Op-amps without stable feedback.....	106
6.4	Problems	111
Chapter 7	Noise.....	115
7.1	Quantifying noise.....	115
7.2	Johnson-Nyquist noise.....	116
7.3	Shot noise.....	118
7.4	Amplifiers	119
7.5	Other noise sources.....	121
7.6	Special topic: digital filtering.....	121
7.7	Problems	123
Chapter 8	Digital Circuits	125
8.1	Binary	125
8.2	Analog-to-digital conversion	126
8.3	Logic gates.....	127
8.4	Digital memory circuits	130
8.5	Clocked digital circuits	134
8.6	Building blocks of a computer.....	138
8.7	Special topic: introduction to quantum computing.....	138
8.7.1	Classical versus quantum bits	138

8.7.2	How to build a qubit.....	140
8.7.3	Quantum gates	141
8.8	Problems	145
Appendix A	Some Useful Constants, Conversions, Identities, Metric Prefixes, and Units.....	147
Appendix B	Review of Complex Numbers	149
Index		151



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Preface

This textbook aims to provide a practical and concise introduction to the subject of electronics for undergraduate students in the physical sciences. The goal is to prepare students to be successful in upper-level laboratory courses and experimental research. The book may also be of interest to anyone wishing to learn or brush up on the fundamentals of electronics. Topics covered include electrical components, circuit theory, signal analysis, and instrumentation. The focus is on analog electronics, but there is a brief introduction to digital electronics in [Chapter 8](#). The book is written assuming that the reader has already completed courses in calculus and introductory electromagnetism, and, for the sake of brevity, some results from electromagnetism are stated without derivation. Sections denoted as *Special topic* can be skipped if needed without compromising the ability to understand subsequent material.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Author

Daniel F. Santavicca is a professor of physics at the University of North Florida, where he runs an active research program in nanoscale electronics with a focus on superconducting thin film devices. He received a PhD in applied physics from Yale University.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

1 Linear DC Circuits

1.1 OHM'S LAW

The electric potential difference, also known as the voltage V , between points x_1 and x_2 is found by integrating the electric field $\vec{E}$ along a path $\vec{l}$ connecting those two points:

$$V = - \int_{x_1}^{x_2} \vec{E} \cdot d\vec{l} \quad (1.1)$$

The negative sign reflects the fact that electric field points from higher electric potential to lower electric potential. The electric force is a conservative force, so the voltage between the two points does not depend on the path taken. Voltage has SI units of volts (V).

If a material experiences an electric field $\vec{E}$, a charge q in the material will feel a force $\vec{F} = q\vec{E}$. If the charge q is positive, $\vec{F}$ and $\vec{E}$ are in the same direction, and if the charge is negative, they are in opposite directions. If the charge q is mobile, then it will move in response to the force. If we apply an electric field along the axial direction of a metal wire, mobile charges will flow along the length of the wire. We define the current I through the wire as the rate at which charge Q passes through one position along the length of the wire:

$$I = \frac{dQ}{dt} \quad (1.2)$$

Current has SI units of amperes (A), often abbreviated as amps. An amp is equal to a coulomb per second (C/s). The direction of the current is, by definition, the direction of positive charge flow. The electron charge is $-q_0$, where $q_0 = 1.6022 \times 10^{-19}$ C is the fundamental charge. Since the electron charge is negative, the direction that mobile electrons move through a metal wire is opposite the direction of the current.

Empirically, we find that the voltage V between opposite ends of a metal wire and the current I flowing through the wire are linearly proportional. We can turn this proportionality into an equality by introducing a constant of proportionality R which we call the resistance:

$$V = IR \quad (1.3)$$

This relationship is called *Ohm's law*. Resistance has SI units of ohms (Ω). The conductance G of a system is simply the inverse of the resistance, $G = 1/R$.

Resistance is a material-dependent property. For a particular material, the resistance also depends on the geometry. It is useful to introduce a quantity that is proportional to the resistance but which is geometry-independent. We call this quantity

the *resistivity* ρ and, for a wire of length L and uniform cross-sectional area A , it is related to the resistance R by:

$$\rho = R \frac{A}{L} \quad (1.4)$$

Resistivity has SI units of ohm meters ($\Omega \text{ m}$). The *conductivity* σ is simply the inverse of the resistivity, $\sigma = 1/\rho$.

The current density $\vec{J}$ is related to the current I by $I = \oint \vec{J} \cdot d\vec{S}$ where the integral is performed over the surface S through which the current flows. For a wire with uniform cross-sectional area A and uniform current density, this simplifies to $I = JA$. Using our definitions of the resistivity and current density, we can re-express Ohm's law in the form $\vec{E} = \rho \vec{J}$ or equivalently:

$$\vec{J} = \sigma \vec{E} \quad (1.5)$$

It may not be obvious why the current and the electric field are linearly proportional. After all, if an electric field causes a force $\vec{F}$ on the mobile charges in a wire, then you would expect, from Newton's second law, $\vec{F} = m\vec{a}$, that the speed of the mobile charges should increase in time, resulting in a current that also increases in time. The reason a DC voltage produces a time-independent current is that the mobile charges experience frequent scattering as they travel through the wire. Each scattering event randomizes the charge's velocity, and the result of this frequent scattering is that the charge moves at some time-independent average velocity known as the *drift velocity*. The magnitude of this drift velocity is much smaller than the magnitude of the velocity at which the mobile charges travel in between scattering events.

One of the first microscope models of electrical current was proposed by Paul Drude in 1900 and is known as the Drude model. This model neglects Coulomb interactions. It assumes that mobile electrons frequently scatter off the fixed ions in a crystal lattice. This assumption is wrong; mobile electrons can travel through a perfect periodic crystal lattice without scattering. Instead, electrons scatter due to defects and impurities in the crystal lattice, the random thermal motion of the ions in the crystal lattice relative to their equilibrium positions, and surfaces. Although his assumption about the mechanism for scattering was incorrect, Drude arrived at useful expressions relating the current density, the drift velocity, and the conductivity to the average time between scattering events τ .

Drude assumed that the average momentum $\langle \vec{p} \rangle$ gained by a mobile charge carrier q in a static electric field $\vec{E}$ in between scattering events is $\langle \vec{p} \rangle = q\vec{E}\tau$. He assumed that each scattering event randomizes the electron's momentum, so the net contribution to the momentum from scattering averages to zero. More specifically, he assumed that the collision randomizes the electron's direction and that the electron's speed following the collision is determined by the local temperature where the scattering event occurred. We can express the average momentum as $\langle \vec{p} \rangle = m\langle \vec{v} \rangle$, where m is the charge carrier mass and $\langle \vec{v} \rangle$ is the average charge carrier velocity. We can write the current density as $\vec{J} = nq\langle \vec{v} \rangle$ where n is the volume density of mobile charges in the material and q is the charge of each mobile charge. Combining these

expressions, we get:

$$\vec{J} = \frac{nq^2\tau}{m}\vec{E} \quad (1.6)$$

Using $\vec{J} = \sigma\vec{E}$, we can write:

$$\sigma = nq^2\tau/m \quad (1.7)$$

We can also combine the previous expressions to write:

$$\langle \vec{v} \rangle = q\tau\vec{E}/m \quad (1.8)$$

We call the magnitude of $\langle \vec{v} \rangle$ the drift velocity v_d . (Even though v_d is a scalar and hence technically a speed, it is still commonly referred to as a velocity.) In general, the scattering time τ is a temperature-dependent quantity, as increasing the temperature will increase the thermal fluctuations of the ions in the crystal lattice relative to their equilibrium positions, increasing the rate at which they scatter electrons.

Exercise 1.1 Applying the Drude Model to Copper

At room temperature, copper has a resistivity $\rho = 1.68 \times 10^{-8} \text{ } \Omega\text{m}$. The density of mobile electrons in copper is $n = 8.49 \times 10^{28} \text{ m}^{-3}$. Find the mean scattering time τ . If there is an electric field whose magnitude is 10 N/C, find the corresponding drift velocity v_d .

Solution: We can express the scattering time in terms of the resistivity as $\tau = m/(nq^2\rho)$. Using the electron mass $m = 9.109 \times 10^{-31} \text{ kg}$, the electron charge $q = -1.602 \times 10^{-19} \text{ C}$, and the given values for n and ρ , we get a scattering time $\tau = 2.49 \times 10^{-14} \text{ s}$. The drift velocity is $v_d = |qE\tau/m|$, so using our value for τ and the given E , we get $v_d = 4.38 \times 10^{-2} \text{ m/s}$.

1.2 INTRODUCTION TO CIRCUIT DIAGRAMS

A circuit component with a resistance given by Ohm's law is called a resistor. We draw resistors in circuit diagrams using the symbol:



Here we chose to draw the resistor symbol horizontally. In a circuit diagram, it can be drawn in any orientation, and the orientation has no relevance to its electrical behavior; this is true of all circuit components.

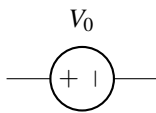
The name of this chapter is *Linear DC Circuits*. *DC* stands for *direct current* and refers to electrical systems in which the currents and voltages are constant in time. A resistor is *linear* if its resistance is independent of the current, i.e., a plot of voltage as

a function of current will be a straight line. For now, we will assume that all resistors are linear, although this assumption will be relaxed when we get to [chapter 5](#).

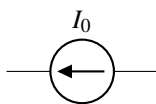
A device that maintains a constant voltage across its terminals is called a DC voltage source. (A *terminal* is a point of connection to a circuit component or a circuit.) One common type of DC voltage source is a battery. Two circuit symbols used to represent a battery with voltage V_0 are:



Technically, the symbol on the left is a single-cell battery, and the symbol on the right is a double-cell battery, although often the two symbols are used interchangeably. In either case, the side with the longer parallel line represents the higher electric potential side (+), and the side with the shorter parallel line represents the lower electric potential side (-); the labels + and - are often omitted for simplicity. Sometimes the battery symbol is used generically to represent any DC voltage source, but we can also use the following symbol for a generic DC voltage source with voltage V_0 :

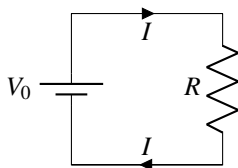


A device that produces a constant current output is called a DC current source. (Of course, no current source can produce any current unless there is a connection between its two terminals.) We use the following circuit symbol for a DC current source that supplies a current I_0 :



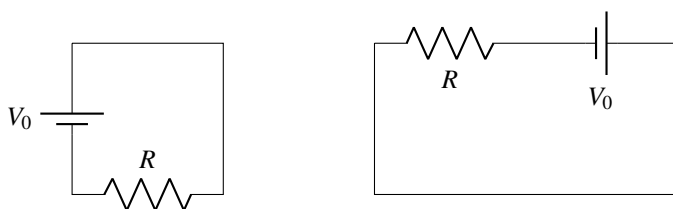
The arrow indicates the direction of current flow. For any current or voltage source, the current flowing out of one terminal is equal to the current flowing into the other terminal.

Now that we have defined several circuit components, we can start connecting them to build some DC circuits. We will start with a simple circuit, a battery connected to a resistor:



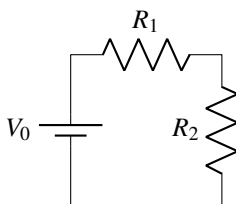
We can use Ohm's law to find the current that flows through the circuit, $I = V_0/R$. Current flows out of the positive terminal of the battery, through the resistor, and then into the negative terminal of the battery. In drawing this circuit, we have surreptitiously introduced a new circuit component: the wire. We represent an ideal wire as a line connecting two different points in the circuit. By *ideal* we mean that the wire has zero resistance, which means, according to Ohm's law, that there is no electric potential difference along the length of the wire. In general, we will assume that wires in circuit diagrams are ideal unless told otherwise.

In circuit diagrams, what is important are the points of connection, not the physical orientation of the components in the diagram. Any two circuit diagrams with all the same points of connection between the components represent the same circuit, even if they are visually different. For example, the following circuits are both electrically the same as the previous circuit:



In all three cases, the positive terminal of the battery connects to one side of the resistor, the negative terminal connects to the other side of the resistor, and the current is $I = V_0/R$.

We can draw a slightly more complicated circuit diagram by adding a second resistor:



In this circuit, the current I will flow out of the positive terminal of the battery, through R_1 , through R_2 , and then back into the negative terminal of the battery. After flowing through R_1 , the current must flow through R_2 ; there is nowhere else for it to go. If two or more circuit components *must* have the same current, then we say that they are *in series*. So here R_1 and R_2 are in series. If N resistors are in series, they have a total equivalent resistance R_{eq} given by:

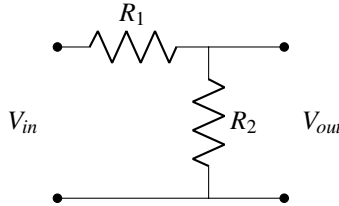
$$R_{eq} = \sum_{n=1}^N R_n \quad (1.9)$$

In the circuit shown above, we have $R_{eq} = R_1 + R_2$. Then we can use this total equivalent resistance in Ohm's law to find the current $I = V_0/R_{eq} = V_0/(R_1 + R_2)$.

The voltage across resistor R_2 , which we will call V_2 , can be found from Ohm's law using the expression for the current that we just found: $V_2 = IR_2 = V_0 R_2 / (R_1 + R_2)$. This situation gives us a circuit that is called a *voltage divider*. If we change the name of V_0 to V_{in} and V_2 to V_{out} , we get an equation for our voltage divider in a more common notation:

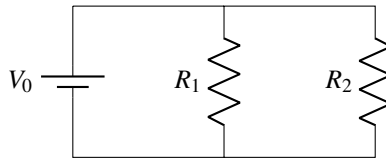
$$V_{out} = V_{in} \frac{R_2}{R_1 + R_2} \quad (1.10)$$

A common way of drawing our voltage divider circuit is as follows:



Here V_{in} is the voltage between the two points indicated by the dots on the left-hand side, and V_{out} is the voltage between the two points indicated by the dots on the right-hand side. Unless told otherwise, we generally assume that V_{in} is provided by an ideal voltage source, and V_{out} is measured with an ideal voltmeter. If, for example, we have $R_1 = R_2$, then $V_{out} = V_{in}/2$, and we would call this a divide-by-two voltage divider. As we progress through this book, we will see the voltage divider circuit makes many appearances, so it is a good idea to remember the voltage divider equation (Equation 1.10).

Another circuit consisting of a battery and two resistors is:



Here both resistors are directly connected to the battery through only ideal wires, so both resistors must have a voltage across them that is equal to the battery voltage V_0 . When we have two or more circuit components that *must* have the same voltage across them, we say that they are *in parallel*. We can find the equivalent resistance R_{eq} of N parallel resistors from:

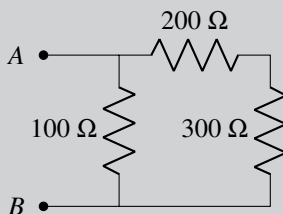
$$R_{eq} = \left(\sum_{n=1}^N \frac{1}{R_n} \right)^{-1} \quad (1.11)$$

Thus, in our example above, the two parallel resistors R_1 and R_2 combine to create an equivalent resistance $R_{eq} = 1/(1/R_1 + 1/R_2)$. The total current I from the battery can then be found from $I = V_0/R_{eq}$. The current flowing through each resistor can be

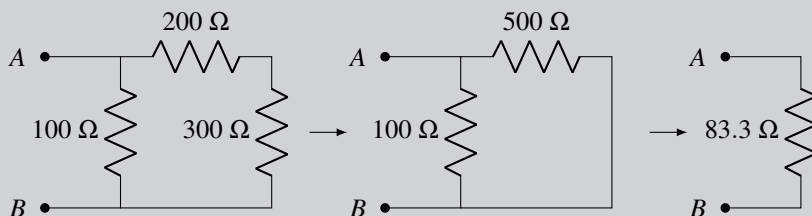
found by applying Ohm's law to each resistor, with the current through R_1 given by $I_1 = V_0/R_1$ and the current through R_2 given by $I_2 = V_0/R_2$. The current that flows out of the positive terminal of the battery has two possible paths it can take back to the negative terminal: it can flow through R_1 , or it can flow through R_2 . Thus we must have $I = I_1 + I_2$. What we have just done is describe the behavior of a *junction*. A junction is a point where multiple electrical components (e.g., wires, resistors, batteries, etc.) connect. In any junction, the total current flowing into the junction must equal the total current flowing out of the junction.

Exercise 1.2 Finding the Equivalent Resistance

Use the rules for combining series and parallel resistors to find the total equivalent resistance between points A and B in the circuit shown below.



Solution: It may be helpful to imagine a current I flowing into the circuit at point A and out of the circuit at point B , resulting in a voltage V between points A and B , with the total equivalent resistance between those two points equal to V/I . We start simplifying the circuit by recognizing that the $200\ \Omega$ and $300\ \Omega$ resistors are in series, so they can be replaced by a single equivalent resistance of $200\ \Omega + 300\ \Omega = 500\ \Omega$. We can then re-draw the equivalent circuit as shown in the center diagram below. This $500\ \Omega$ resistor is now in parallel with the $100\ \Omega$ resistor, resulting in a final equivalent resistance of $1/(1/100\ \Omega + 1/500\ \Omega) = 83.3\ \Omega$, as shown in the right diagram below.



Note that it is not always possible to simplify an arbitrary combination of resistors into a single equivalent resistance by combining series and parallel resistors. We will soon learn other techniques for analyzing such circuits.

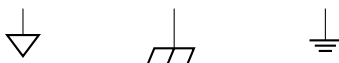
A component related to the resistor is the potentiometer, sometimes called a *pot* for short. The circuit symbol for a potentiometer is:



The potentiometer has three points of connection corresponding to the left and right horizontal wires and the top vertical wire in the diagram above. The left and right points connect across a fixed resistance, while the top connection – known as the *wiper* – can be physically moved along the length of the resistor with a knob. The resistance measured between the wiper and either the left or right point varies as one turns the knob, creating a variable resistor. (Such a variable resistor is sometimes called a *rheostat*.) The sum of the resistance between the wiper and the left point and the resistance between the wiper and the right point equals the total fixed resistance.

Ground refers to a point in the circuit where we have defined the electric potential to be zero. One always has the freedom to choose where to define the zero of potential energy in a system, so we can choose any point in a circuit and define that as ground. Then the electric potential at any other point in the circuit is defined with respect to the point we specified as ground. For example, if we have a 9 V battery and we call the positive terminal ground, then we are defining that as 0 V, which means that the negative terminal is at a potential of -9 V. If instead we call the negative terminal ground, that becomes defined as 0 V and hence the positive terminal must be at a potential of $+9$ V. In either case, the battery is doing its job by maintaining the positive terminal 9 V higher in electric potential than the negative terminal.

If we have multiple points in a circuit diagram defined as ground, then those points behave as if they are connected, since they are tied to the same potential. (In an actual circuit, any points that are ground should be directly connected to ensure that they always remain at the same potential.) Ground can be represented in circuit diagrams with one of the following symbols:



The left symbol represents signal ground, which is a ground that is defined locally within a circuit. Two different circuits may each have their own different definitions of signal ground. The middle symbol represents chassis ground; this means the ground of the circuit is connected to the metal enclosure in which it is housed. This is done for safety reasons, to prevent a loose wire from accidentally charging up the enclosure to a high potential and shocking the user. The right symbol represents earth ground, which is a ground that is connected to a metal rod or pipe that goes into the earth. Often this means that the circuit ground is connected to the third prong of its power cord, with the assumption that the power line ground is an earth ground.

1.3 POWER IN DC CIRCUITS

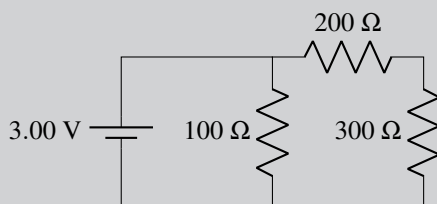
If a circuit or circuit component carries a DC current I and has a DC voltage V across it, the power P dissipated in that circuit or circuit component is given by:

$$P = IV \quad (1.12)$$

Power has SI units of watts (W), where it is useful to remember that a watt is equal to a joule per second (J/s). Using Ohm's law, we can write Equation 1.12 equivalently as $P = I^2 R$ or $P = V^2/R$, where R is the resistance of the circuit or circuit component. A component must have some nonzero resistance in order to dissipate power. Typically, power dissipated in a resistor represents the conversion of electrical energy into thermal energy, i.e., the resistor heats up. The electrical energy is supplied by the circuit's voltage or current source.

Exercise 1.3 Power in DC Circuits

Here we will consider the same resistor circuit from [exercise 1.2](#), but now we will connect a 3.00 V battery between terminals A and B , as shown below. First, find the total power dissipated in the circuit. Then, find the power dissipated in the 200 Ω resistor.



Solution: We can use the result of [exercise 1.2](#), which found that the total equivalent resistance of the circuit is $R_{eq} = 83.3 \Omega$, to find the total power dissipated in the circuit using $P_{tot} = V^2/R_{eq} = (3.00 \text{ V})^2/83.3 \Omega = 0.108 \text{ W}$. To find the power dissipated in the 200 Ω resistor, we need to find either the current through that resistor or the voltage across that resistor. One way to do this is to recognize that the series combination of the 200 Ω and 300 Ω resistors are connected directly across the 3.00 V source, so we can find the current through these series resistors using their equivalent resistance of 500 Ω in Ohm's law: $I = V/R = 3.00 \text{ V}/500 \Omega = 6.00 \text{ mA}$. Since series resistors have the same current, this is the current through the 200 Ω resistor. So now we can find the power dissipated in the 200 Ω resistor using $P_{200\Omega} = I^2 R = (6.00 \times 10^{-3} \text{ A})^2 200 \Omega = 7.20 \text{ mW}$.

1.4 KIRCHHOFF'S RULES

Kirchhoff's rules provide a systematic approach to solving more complicated circuit problems. They were first developed by the German physicist Gustav Kirchhoff in 1845 while he was still a student. There are two rules, often called the loop rule and the junction rule:

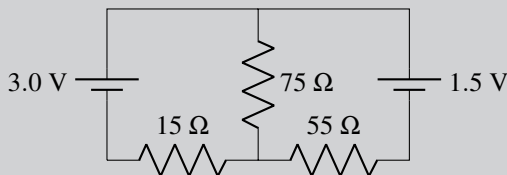
- (1) **Loop Rule:** The sum of the voltage changes around any closed loop in a circuit is equal to zero.
- (2) **Junction Rule:** The sum of the current flowing into a junction is equal to the sum of the current flowing out of the junction.

Another way to state the loop rule is that, if we follow some continuous path through a circuit that leads back to where we started, the total change in electric potential must be zero. To apply the loop rule, it can be helpful to draw arrows labeling the direction of current flow through each resistor. Remembering that current always flows from the higher potential side of a resistor to the lower potential side will help us establish a consistent sign convention. When we go across a resistor R in the same direction as the current I , we are going down in electric potential and will write that as a negative voltage change equal to $-IR$. When we go across a resistor R in the opposite direction of the current I , we are going up in electric potential and will write that as a positive voltage change equal to IR . If we are not sure which direction the current is flowing through a particular resistor, we can just guess and draw the arrow based on our guess. If we guess incorrectly, we will get a negative value for that current, but the value will still be correct and we can still use it (making sure to keep it as a negative number) to solve for the other circuit parameters.

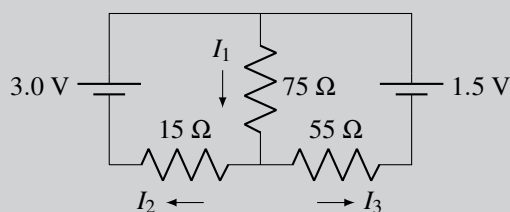
In general, we want to use Kirchhoff's rules to write as many independent equations as there are unknowns. Then we can solve the system of equations for the unknowns. Probably the best way to show how to do this is through an example.

Exercise 1.4 Applying Kirchhoff's Rules

Use Kirchhoff's rules to find the current through each of the three resistors in the circuit shown.



Solution: We'll start by drawing current arrows through each of our three resistors, taking our best guess for the direction of each one:



Since we have three unknowns, we want to write three independent equations. We will do this using two loop rules and one junction rule. (There are only two nontrivial junctions in this circuit, and they will give equivalent junction rule equations.) It is good to remember that the current flowing out of one terminal of a voltage source must be equal to the current flowing into the other terminal.

For the left loop, we will start at the low-potential side of the 3.0 V battery and go around the loop clockwise. (Note that we are able to choose any starting point and any direction of travel around the loop.) When we go across the 3.0 V battery, we are going from the low-potential side to the high potential side, so the voltage change is +3.0 V. Continuing clockwise, we go across an ideal wire, which has no voltage change. Then we go across the 75 Ω resistor in the direction of current flow, so the voltage change is $-(I_1 75 \Omega)$. Then we go across the 15 Ω resistor in the direction of the current flow, so the voltage change is $-(I_2 15 \Omega)$. Now we're back to our starting point, so we add up our voltage changes and set them equal to zero:

$$3.0 - 75I_1 - 15I_2 = 0$$

For simplicity, we have left off the units, but we know that, since all quantities are given in base SI units (e.g., Ω instead of $\text{k}\Omega$), our answers must be in the appropriate base SI unit.

We can now take the same approach for the right loop. This time we'll start at the low-potential side of the 1.5 V battery and go counterclockwise around the loop, resulting in the equation:

$$1.5 - 75I_1 - 55I_3 = 0$$

Finally, we will write our junction rule for the bottom junction (where the three resistors are connected):

$$I_1 = I_2 + I_3$$

Now, our task is to solve this system of three equations for the three unknowns. We can do this via substitution. Our first step will be to rewrite the junction rule as $I_3 = I_1 - I_2$ and then plug this into our right loop equation, which gives us $1.5 - 75I_1 - 55(I_1 - I_2) = 1.5 - 130I_1 + 55I_2 = 0$. Solving this for I_2 , we get $I_2 = (130/55)I_1 - 1.5/55$. We can then plug this into our left loop equation, which gives us $3.0 - 75I_1 - 15((130/55)I_1 - 1.5/55) = 0$. Solving for I_1 , we

get $I_1 = 0.03086$ A. We can plug this back into our previous expression for I_2 , which gives us $I_2 = 0.04568$ A. Then, we can plug these into the junction rule equation to get $I_3 = -0.01481$ A. The negative sign on I_3 means that we drew our current arrow in the wrong direction, but the value is still correct. We see that there is current flowing out of the negative terminal of the 1.5 V battery and into its positive terminal, which is not normally how a voltage source operates. This is the result of the larger 3.0 V battery. In a real circuit, we would probably want to avoid a design that would do this (unless we are trying to recharge a rechargeable battery). Now that we have found our final answers, we can round to the number of significant figures (sigfigs) that were used in the values given in the original problem, i.e., $I_1 = 0.031$ A, $I_2 = 0.046$ A, and $I_3 = -0.015$ A. Note that we kept extra sigfigs in the intermediate steps and waited until the end to round our values to two sigfigs; rounding too much in intermediate steps can introduce rounding error.

For those who are familiar with linear algebra, we can also represent our original three Kirchhoff's rules equations as a matrix equation. We will have a 3×3 matrix in which each row represents the factors multiplying I_1 , I_2 , and I_3 , in that order, in each of our three equations. This matrix is multiplied by the current vector to get the resulting vector consisting of voltages for our two loop equations and zero for our junction equation:

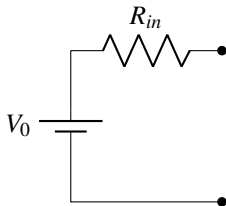
$$\begin{bmatrix} 75 & 15 & 0 \\ 75 & 0 & 55 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 3.0 \\ 1.5 \\ 0 \end{bmatrix}$$

We can then invert the 3×3 matrix to solve for the currents:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 75 & 15 & 0 \\ 75 & 0 & 55 \\ 1 & -1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 3.0 \\ 1.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.031 \\ 0.046 \\ -0.015 \end{bmatrix}$$

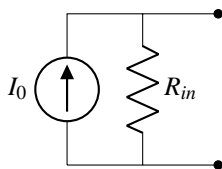
1.5 SOURCES AND METERS

If a voltage source such as a battery with voltage V_0 is connected across a resistor with resistance R , the resulting current is $I = V_0/R$. In stating this, we have assumed that the voltage source itself has no resistance. An *ideal* voltage source is one that has zero internal resistance. Real voltage sources have some nonzero internal resistance, although in many cases, it is small enough that we can approximate them as ideal. If we see a voltage source in a circuit diagram, we usually assume that it can be treated as ideal. We can represent a nonideal voltage source in a circuit diagram as an ideal voltage source V_0 in series with a separate resistor R_{in} that represents the internal resistance of the nonideal source:



If we connect an external resistor R across this nonideal voltage source, then R and R_{in} are in series, and they have a total equivalent resistance $R_{eq} = R + R_{in}$. The resulting current will be $I = V_0/R_{eq} = V_0/(R + R_{in})$. If $R_{in} \ll R$, then $I \approx V_0/R$ and we approximate the ideal case.

An ideal current source has infinite internal resistance. (We can imagine creating an ideal current source by placing an infinite resistor in series with a voltage source. Of course, if the resistance is infinite, the resulting current will be zero. Hence it is not possible to use this approach to make a current source that is actually ideal.) A real current source has some finite but typically very large internal resistance. We can model a nonideal current source as an ideal current source I_0 in parallel with a separate resistor R_{in} that represents the internal resistance of the nonideal source:



If we connect an external resistor R across our nonideal current source, it will be in parallel with R_{in} , resulting in a total equivalent resistance of $R_{eq} = 1/(1/R + 1/R_{in})$. The voltage across these parallel resistors will be $V = I_0 R_{eq}$ and the current through the external resistor will be $I = V/R = I_0 R_{in}/(R + R_{in})$. If $R_{in} \gg R$, then $I \approx I_0$ and we approximate the ideal case.

A voltmeter measures the voltage between two points in a circuit. A meter has a high potential side and a low-potential side (sometimes referred to as ground). If we connect the high side of the voltmeter to the higher electric potential point in the circuit, and the low side of the voltmeter to the lower electric potential point in the circuit, the voltmeter will report a positive voltage. If the connections are reversed, the voltmeter will report a negative voltage. A circuit symbol for a voltmeter is:



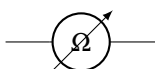
The arrow alludes to the indicator needle on a meter with an analog display, although today it is more common to encounter digital meters with numeric displays.

An ammeter measures the current flowing through the meter. If the current flows into the high potential side and out of the low-potential side of the meter, the ammeter

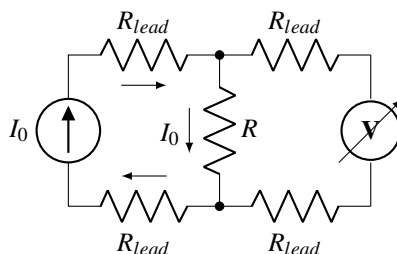
will report it as a positive current. If the connections are reversed, the ammeter will report it as a negative current. A circuit symbol for an ammeter is:



A multimeter is a device that can measure multiple physical quantities. You may encounter the acronym DMM, which stands for digital multimeter. Most DMMs can function as a voltmeter, and ammeter, and an ohmmeter. (Some can also measure other quantities such as frequency, capacitance, or inductance.) An ohmmeter is really just a current source combined with a voltmeter. The ohmmeter produces a known current, measures the resulting voltage, and calculates the resistance from Ohm's law. When using an ohmmeter, it is important that there be no other sources of current present in the circuit, as that can cause an incorrect resistance reading. A circuit symbol for an ohmmeter is:

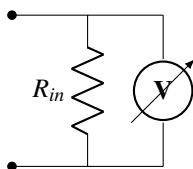


If we connect an ohmmeter to some component with resistance R , we actually measure a resistance $R + 2R_{lead}$, where R_{lead} is the resistance of each of the two leads (wires) used to connect the ohmmeter to the component. If $R_{lead} \ll R$, then we may not need to worry about this. But, if needed, we can eliminate the effect of the lead resistance using a technique called a *four-wire measurement*, as shown here:



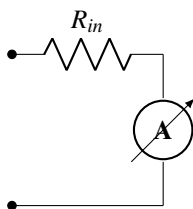
In this configuration, a current source is connected to component R using one pair of wires and a voltmeter is connected to the component using a separate pair of wires. There is no current flowing through the wires connecting to the voltmeter and hence there is no voltage drop across these wires. This means that the ratio of the measured voltage to the source current will yield the true component resistance R .

An ideal meter should not change the system being measured, i.e., the system should behave the same both with and without the meter connected. For a voltmeter to be ideal, it must have infinite internal resistance. (In our discussion of the four-wire measurement, we assumed that the voltmeter was ideal.) A nonideal (real) voltmeter has some finite but typically very large internal resistance. We can model a nonideal voltmeter as an ideal voltmeter in parallel with a separate resistor R_{in} that represents its internal resistance:



A voltmeter should be connected in parallel with the component whose voltage one is measuring. If a nonideal voltmeter is connected in parallel with a component with resistance R , it will behave as approximately ideal provided that $R_{in} \gg R$. In [Exercise 1.5](#), we will show an example of calculating the effect of a nonideal voltmeter on the measured voltage.

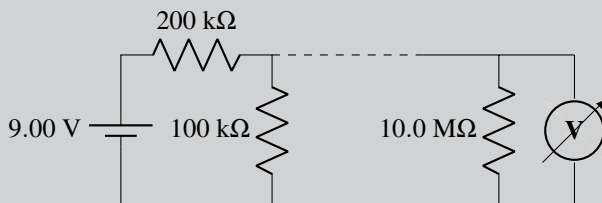
An ammeter should be connected in series with the component whose current one is measuring. In order not to change the current through the component, an ideal ammeter must have zero internal resistance. A nonideal (real) ammeter has some nonzero but typically very small internal resistance. We can model a nonideal ammeter as an ideal ammeter in series with a separate resistor R_{in} representing the ammeter's internal resistance:



If a nonideal ammeter is connected in series with a circuit with equivalent resistance R , the ammeter will behave as approximately ideal as long as $R_{in} \ll R$

Exercise 1.5 Effect of a Nonideal Voltmeter

A nonideal voltmeter with internal resistance $R_{in} = 10.0 \text{ M}\Omega$ is connected in parallel with the $100 \text{ k}\Omega$ resistor in the circuit shown. What voltage will the voltmeter measure? And what is the voltage across the $100 \text{ k}\Omega$ resistor when the voltmeter is not connected?

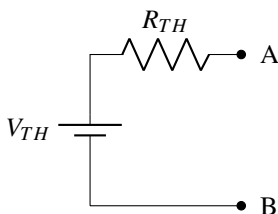


Solution: This circuit is a voltage divider, so we know that, without the voltmeter connected, the voltage across the $100 \text{ k}\Omega$ resistor is $V = 9.00 \times 100 / (200 +$

100) = 3.00 V. When we connect the voltmeter in parallel with the 100 k Ω resistor, its internal resistance of 10.0 M Ω is in parallel with the 100 k Ω resistor, resulting in an equivalent resistance $R_{eq} = 1/(1/100 \text{ k}\Omega + 1/10.0 \text{ M}\Omega) = 99.0 \text{ k}\Omega$. We can now use this new value in our voltage divider formula to find the voltage measured by the voltmeter: $V = 9.00 \times 99.0/(200 + 99.0) = 2.98 \text{ V}$. We see the finite input resistance of the voltmeter results in the measured voltage being 0.02 V lower than the “true” voltage (the voltage without the voltmeter connected).

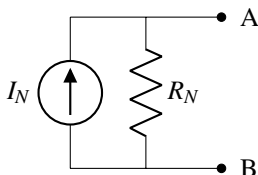
1.6 THEVENIN AND NORTON EQUIVALENT CIRCUITS

If one defines two points of connection to a circuit consisting of any combination of resistors, voltage sources, and current sources, it is always possible to describe the behavior in terms of either a Thevenin or Norton equivalent circuit. The definition of the Thevenin or Norton equivalent circuit is specific to the two points of connection, so different points of connection to the same circuit will, in general, yield a different equivalent circuit. The model of the Thevenin equivalent circuit consists of an ideal DC voltage source V_{TH} in series with a resistor R_{TH} , where we have labeled the two points of connection *A* and *B*:



Note that this diagram applies to *any* Thevenin equivalent circuit; only the values of the Thevenin voltage V_{TH} and the Thevenin resistance R_{TH} will vary. Finding the values of V_{TH} and R_{TH} is a straightforward process. First one connects an ideal voltmeter to the two specified points in the circuit; the voltage measured by the voltmeter is V_{TH} . Then one connects an ideal ammeter to the same two points; we will call the current measured by the ideal ammeter the short-circuit current I_S . Then R_{TH} can be found from $R_{TH} = V_{TH}/I_S$.

The Norton equivalent circuit consists of an ideal DC current source I_N in parallel with a resistor R_N , where again, we have labeled the two points of connection *A* and *B*:

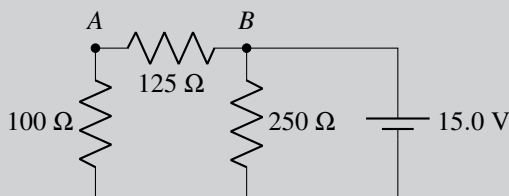


This diagram applies to *any* Norton equivalent circuit. To find the Norton current I_N and the Norton resistance R_N , one performs the same two measurements as for finding the Thevenin equivalent circuit: one connects an ideal voltmeter to the two points of connection and then connects an ideal ammeter to the same two points of connection. Here the measured current is I_N and we will call the measured voltage the open-circuit voltage V_O . The Norton resistance can then be found from $R_N = V_O/I_N$. The Norton resistance is always the same value as the Thevenin resistance.¹

Defining a Thevenin or Norton equivalent circuit can be useful because it often simplifies the task of determining what happens if one connects some other circuit or circuit component to the two points in the original circuit. For example, if we have some complex circuit for which we know its Thevenin equivalent circuit relative to two points A and B, and we connect an external resistor R_L between those two points, then finding the voltage V_L across the external resistor becomes simple. This is because the Thevenin equivalent circuit with R_L connected between A and B is just a voltage divider, so $V_L = V_{TH}R_L/(R_L + R_{TH})$.

Exercise 1.6 Finding the Thevenin Equivalent Circuit

Find the Thevenin equivalent circuit (i.e., find V_{TH} and R_{TH}) for the circuit shown for the two points of connection A and B.

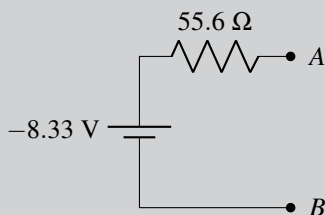


Solution: First, we imagine connecting an ideal voltmeter to points A and B. No current flows through our ideal voltmeter (because it has infinite resistance), so the 100 Ω resistor is in series with the 125 Ω resistor, resulting in a total

¹The Thevenin/Norton resistance can also be found by replacing any voltage sources in the circuit with wires (because an ideal voltage source has zero internal resistance) and replacing any current sources with open circuits (because an ideal current source has infinite internal resistance) and then finding the equivalent resistance between the two points of connection. This equivalent resistance is then the Thevenin resistance R_{TH} or Norton resistance R_N .

equivalent resistance of $225\ \Omega$. This equivalent resistance is directly connected to both sides of the battery, so we can find the current through the equivalent resistance from Ohm's law: $(15.0\ \text{V})/(225\ \Omega) = 0.06667\ \text{A}$. This is also the current through the $125\ \Omega$ resistor, which means that the voltage across the $125\ \Omega$ resistor is $(0.06667\ \text{A})(125\ \Omega) = 8.33\ \text{V}$. Since the current flows from point B to point A , point B must be higher in electric potential. Based on the drawing of the Thevenin equivalent circuit, this means that V_{TH} is negative, i.e., $V_{TH} = -8.33\ \text{V}$.

Next, we imagine connecting an ideal ammeter to the same two points. The ammeter is in parallel with the $125\ \Omega$ resistor, and since an ideal ammeter has zero resistance, the equivalent resistance of the parallel combination of the ammeter and the resistor is zero. This equivalent resistance is in series with the $100\ \Omega$ resistor, resulting in an equivalent series resistance of $100\ \Omega$. The current through this equivalent resistance is $(15.0\ \text{V})/(100\ \Omega) = 0.150\ \text{A}$. This current will flow out of the ammeter at point A (the high potential side) and into the ammeter at point B (the low-potential side), so we call it a negative current, i.e. $I_S = -0.150\ \text{A}$. Then we can find R_{TH} from $R_{TH} = V_{TH}/I_S = (-8.33\ \text{V})/(-0.150\ \text{A}) = 55.6\ \Omega$. The Thevenin equivalent circuit always has the same form, so it isn't really necessary to draw it, but if we want to, we can:



The negative sign on the voltage source means that point A is lower in electric potential than point B , consistent with the original circuit. (If we prefer, we could switch the orientation of the battery and drop the negative sign.)

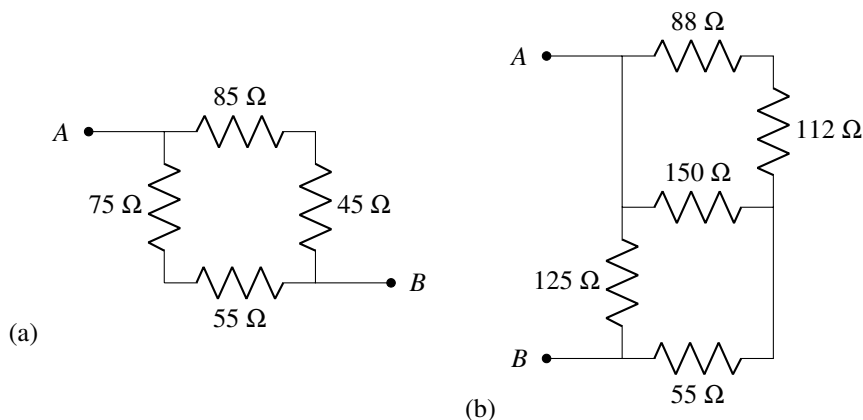
1.7 PROBLEMS

Problem 1. A $9.00\ \text{V}$ battery is connected across a $50.0\ \text{k}\Omega$ resistor. How many electrons flow through the resistor every second?

Problem 2. The energy storage capacity of batteries is often specified in units of milliamp hours (mAh) or, for larger batteries, amp hours (Ah). For example, a battery specified at $1,500\ \text{mAh}$ should be capable of supplying a current of $1,500\ \text{mA}$ for a total duration of 1 hour. If a $1.5\ \text{V}$ battery is rated for $1,000\ \text{mAh}$, how much energy is stored in the battery when it is new?

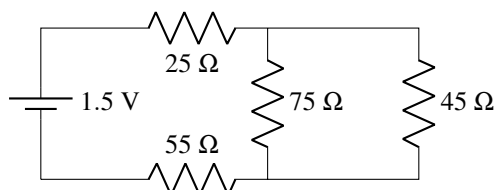
Problem 3. The metal niobium has a mobile electron density $n = 8.56 \times 10^{28} \text{ m}^{-3}$ and a room-temperature resistivity $\rho = 1.52 \times 10^{-7} \Omega\text{m}$. If a room-temperature niobium wire experiences an electric field with a magnitude of 5.50 N/C along its length, use the Drude model to find (a) the average electron scattering time τ , (b) the drift velocity v_d , and (c) the magnitude of the current density J .

Problem 4. Use the rules for combining series and parallel resistors to find the total equivalent resistance between points A and B in each of the two circuits shown below. (In other words, what resistance would you measure if you connected an ohmmeter to points A and B ?)



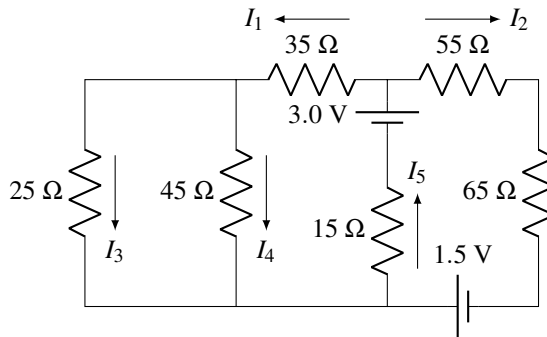
Problem 5. Draw a circuit containing a battery and four resistors R_1 , R_2 , R_3 , and R_4 (a) in which R_1 and R_2 are in series and R_3 and R_4 are in parallel, and (b) in which R_1 and R_2 are in parallel and R_3 and R_4 are not in series or in parallel with any other individual resistor. Your circuits should be such that some nonzero current flows through each resistor.

Problem 6. For the circuit shown below, find (a) the total power dissipated in the circuit and (b) the power dissipated in the 75Ω resistor.

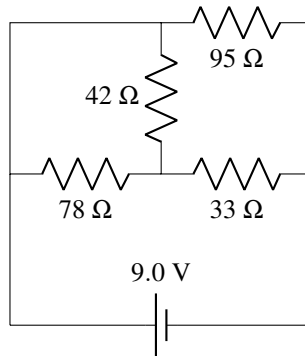


Problem 7. A 120Ω resistor and a 180Ω resistor are connected in series. Each resistor will burn out if the power dissipated by that resistor reaches 0.250 W . You connect an adjustable DC voltage source across the two series resistors. Starting at zero volts, you gradually turn up the source voltage until one of the resistors burns out. At what source voltage does the first resistor burn out, and which resistor will burn out first?

Problem 8. For the circuit shown below, use Kirchhoff's rules to write five independent equations containing the five unknowns I_1 through I_5 . Do not introduce any additional unknowns. You do not need to solve for the five currents; just write five independent equations.

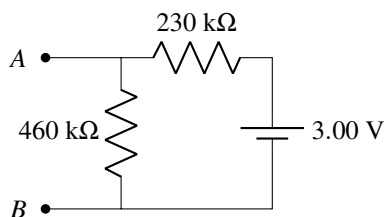


Problem 9. In the circuit shown below, find the current through each resistor. Call the current through $78\ \Omega$ resistor I_1 , the current through the $33\ \Omega$ resistor I_2 , the current through the $42\ \Omega$ resistor I_3 , and the current through the $95\ \Omega$ resistor I_4 . Then find the total power dissipated in the circuit.

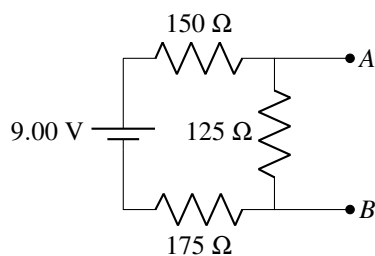


Problem 10. A $1.50\ \text{V}$ DC voltage source is connected across a $4.50\ \Omega$ resistor. What is the current through the resistor if (a) the voltage source is ideal, (b) the voltage source has an internal resistance of $0.100\ \Omega$, and (c) the voltage source has an internal resistance of $1.00\ \Omega$?

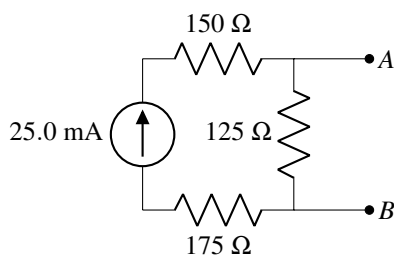
Problem 11. What is the voltage between points A and B measured with (a) an ideal voltmeter, (b) a voltmeter with an internal resistance of $10.0\text{ M}\Omega$, and (c) a voltmeter with an internal resistance of $1.00\text{ M}\Omega$?



Problem 12. Find both the Thevenin equivalent circuit and the Norton equivalent circuit seen at points A and B .



Problem 13. Find both the Thevenin equivalent circuit and the Norton equivalent circuit seen at points A and B .





Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

2 Linear AC Circuits

AC stands for *alternating current* and refers to circuits whose currents and voltages are not constant in time. When we talk about AC circuits, we often assume that the currents and voltages oscillate as sinusoidal functions of time, although this is not always the case. A *linear* AC circuit component is one whose properties do not depend on the amplitude of the current through the component. In this chapter, we will assume that our components are linear. This assumption will be relaxed in later chapters.

2.1 CAPACITANCE AND INDUCTANCE

If we have two oppositely-charged conductors that are not in contact that each contain a net charge magnitude Q , there will be an electric field pointing from the positively-charged conductor to the negatively charged conductor. If we integrate the electric field along some path that goes from the surface of one conductor to the surface of the other conductor, we will find the voltage V between the two. We find that Q and V are linearly proportional, and so we introduce a constant of proportionality that we call the capacitance C :

$$Q = CV \quad (2.1)$$

This is our defining equation for capacitance, in the same way, that Ohm's law is our defining equation for resistance. Capacitance has SI units of farads (F).

Equation 2.1 is true for any capacitor, regardless of geometry. For a particular geometry known as a parallel-plate capacitor, we can arrive at a convenient expression for the capacitance in terms of the physical properties of the capacitor. A parallel plate capacitor has two identical metal plates, each with surface area A , separated by a uniform distance d . Assuming that d is very small compared to the length and width of each plate, we find that the capacitance is given by

$$C = \frac{\kappa \epsilon_0 A}{d} \quad (2.2)$$

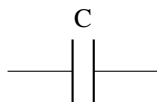
where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m is a constant known as the permittivity of free space and κ is the relative permittivity – also known as the dielectric constant – of the medium between the plates. Equations 2.1 and 2.2 can be derived using Gauss's law; these derivations are found in most introductory physics textbooks. κ is a unitless number that is greater than or equal to one for conventional materials and quantifies the electric polarizability of the material. For free space, $\kappa = 1$, and for other insulating materials, $\kappa > 1$. For dry air at atmospheric pressure and room temperature,

$\kappa = 1.00059$, which is close enough to the free space value that we often treat free space and air as equivalent.

A component that exhibits capacitance is called a capacitor. If we recall that our expression for power (Equation 1.12) is $P = IV$, and we plug in the voltage across a capacitor $V = Q/C$, we get $P = IQ/C$. Using $I = dQ/dt$, we can write $P = (dQ/dt)(Q/C)$. Energy is found by integrating power over time, so the energy stored in a capacitor can be given by $E = \int P dt = \int (dQ/dt)(Q/C) dt = (1/C) \int Q dQ$. If we integrate from zero to Q , we get $E = Q^2/(2C)$. Substituting $Q = CV$, we arrive at the following expression for the energy stored in a capacitor:

$$E = \frac{1}{2} CV^2 \quad (2.3)$$

The circuit symbol for a capacitor with capacitance C is:



Series and parallel capacitors can be combined into a single equivalent capacitance, but they do so in the opposite way that series and parallel resistors combine. If we have N series capacitors, then the total equivalent capacitance C_{eq} is given by

$$C_{eq} = \left(\sum_{n=1}^N \frac{1}{C_n} \right)^{-1}. \quad (2.4)$$

Note that series capacitors must each have the same charge Q . If we have N parallel capacitors, the total equivalent capacitance is

$$C_{eq} = \sum_{n=1}^N C_n. \quad (2.5)$$

And, of course, each parallel capacitor must have the same voltage V .

At zero frequency (DC), a capacitor is simply an open circuit (which is another way of saying it has infinite resistance), as the two sides of a capacitor are not in contact. But at finite frequency, a capacitor is more interesting, as we will see in the next section.

If we pass a current through a wire, it creates a magnetic field. Changing the magnetic field induces a voltage that opposes this change. (This is known as Lenz's law.) We find that the induced voltage V is proportional to the time rate of change of the current dI/dt . We can then introduce a constant of proportionality that we call the self-inductance L , often abbreviated as just inductance:

$$V = L \frac{dI}{dt} \quad (2.6)$$

This is our defining equation for inductance. Inductance has SI units of henries (H).

Equation 2.6 is valid for any inductor. For a particular inductor geometry called a solenoid, we can find a useful expression for the inductance in terms of the physical parameters. A solenoid consists of a wire wound into N loops, with each loop wound in the same direction and having the same loop area A . The loops are stacked one next to the other along a total length ℓ (measured along the solenoid's axial direction). The total inductance of our solenoid is

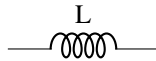
$$L = \frac{\mu_r \mu_0 N^2 A}{\ell} \quad (2.7)$$

where $\mu_0 = 4\pi \times 10^{-7}$ H/m $\approx 1.2566 \times 10^{-6}$ H/m is a constant known as the permeability of free space and μ_r is a unitless number known as the relative permeability. For nonmagnetic materials, $\mu_r = 1$. Paramagnetic and ferromagnetic materials have $\mu_r > 1$, and diamagnetic materials have $\mu_r < 1$. Derivations of Equations 2.6 and 2.7 using Faraday's law can be found in most introductory physics textbooks.

A circuit component that exhibits inductance is called an inductor. If we plug Equation 2.6 into our expression for power, $P = IV$, we get $P = IL(dI/dt)$. We can then multiply both sides by dt . Integrating the left-hand side gives the energy, $\int P dt = E$. Integrating the right-hand side from zero to I gives $L \int I dI = (1/2)LI^2$. Hence the energy stored in an inductor L that carries a current I is given by:

$$E = \frac{1}{2}LI^2 \quad (2.8)$$

The circuit symbol for an inductor with inductance L is:



Series and parallel inductors combine in the same way as series and parallel resistors. If we have N series inductors, then the total equivalent inductance L_{eq} is given by

$$L_{eq} = \sum_{n=1}^N L_n. \quad (2.9)$$

And if we have N parallel inductors, the total equivalent inductance is

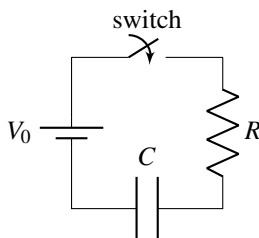
$$L_{eq} = \left(\sum_{n=1}^N \frac{1}{L_n} \right)^{-1}. \quad (2.10)$$

At zero frequency (DC), an ideal inductor is simply a short circuit (which is another way of saying it has zero resistance), since it is just a wire. A real inductor has some resistance, but often it is low enough that we can neglect it and treat the inductor as ideal. We can model a nonideal inductor as an ideal inductor in series with a resistor whose value is equal to the resistance of our nonideal inductor.

2.2 TRANSIENT ANALYSIS

2.2.1 RC CIRCUIT

Here we will consider circuits with a DC voltage source in which we create time-dependent behavior by closing a switch. This will result in a transient time-dependent response which will eventually die out, leaving just the steady-state (DC) behavior. First we will consider a circuit consisting of a series combination of a resistor R and a capacitor C , known as a series RC circuit, that is connected via a switch to a DC voltage source V_0 :



Since everything is in series, the ordering of the components in the circuit does not matter; as long as everything remains in series, the behavior will be the same.

We will assume that the switch is initially open and the capacitor is initially uncharged, and we will determine what happens after closing the switch at time $t = 0$. With the switch open ($t < 0$), no current can flow through the circuit, and there is no voltage across the resistor or the capacitor. Once the switch closes, a time-dependent current $I(t)$ will flow. After the transient response dies out (in the limit $t \rightarrow \infty$), we expect $I(\infty) = 0$ because the capacitor is an open-circuit at DC. In the $t \rightarrow \infty$ limit, we also expect that the voltage across the capacitor will be equal to the source voltage V_0 , which means that the capacitor must have a charge $Q = V_0 C$. The transient response is what allows the capacitor to go from being uncharged at $t \leq 0$ to having charge $Q = V_0 C$ at $t \rightarrow \infty$.

The time-dependent voltage across the resistor can be written using Ohm's law as $V_R(t) = I(t)R$. The time-dependent voltage across the capacitor can be written using Equation 2.1 as $V_C(t) = Q(t)/C$ where $Q(t)$ is the time-dependent capacitor charge. We can then write a loop rule equation for $t \geq 0$ (with the switch closed): $0 = V_0 - V_R(t) - V_C(t) = V_0 - I(t)R - Q(t)/C$. We wish to recast this equation in terms of a single time-dependent variable. Using $I(t) = dQ/dt$, we can write the loop equation as:

$$\frac{dQ}{dt} = \frac{V_0}{R} - \frac{Q(t)}{RC}$$

This is a first-order differential equation for $Q(t)$. We see that we need a solution for $Q(t)$ whose derivative gives us back a constant plus the original function multiplied by a constant. This naturally leads us to an exponential solution of the general form $Q(t) = A + Be^{-t/\tau}$ where A , B , and τ are placeholders for constants that we must

determine in order to find our final expression for $Q(t)$.¹ This solution will be valid for $t \geq 0$. We know that τ must have units of time, since the argument of the exponential must be unitless. τ is what we call the exponential time constant, since $e^{-t/\tau}$ falls to $e^{-1} \approx 0.37$ after a time $t = \tau$.

Since the capacitor is initially uncharged, we must have $Q(0) = 0$. Inserting this into our solution, we get $A + B = 0$ or $A = -B$. Using this, we can write our solution as $Q(t) = A(1 - e^{-t/\tau})$. When the transient response dies out ($t \rightarrow \infty$), we must have $Q(\infty) = V_0C$. Plugging this into our solution, we get $A = V_0C$. All that remains now is to find an expression for τ . To do this, we will plug our solution back into the original differential equation. Plugging in our solution $Q(t) = V_0C(1 - e^{-t/\tau})$ and its derivative $dQ/dt = (V_0C/\tau)e^{-t/\tau}$ and solving for τ , we get $\tau = RC$. (One can verify via unit analysis that an ohm times a farad equals a second.) We now have our final expression for the time-dependent charge on the capacitor:

$$Q(t) = V_0C \left(1 - e^{-t/(RC)}\right)$$

Plugging this into Equation 2.1 gives the expression for the time-dependent voltage across the capacitor V_C :

$$V_C(t) = V_0 \left(1 - e^{-t/(RC)}\right)$$

We can use the loop rule, $V_0 - V_R(t) - V_C(t) = 0$, to find the time-dependent voltage across the resistor V_R :

$$V_R(t) = V_0 e^{-t/(RC)}$$

Plugging $V_R(t)$ into Ohm's law gives the time-dependent current through the resistor $I(t)$:

$$I(t) = (V_0/R) e^{-t/(RC)}$$

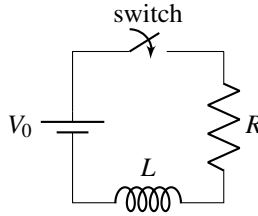
The resistor and capacitor are in series, so they have the same current. Hence we could also find $I(t)$ by taking the time-derivative of $Q(t)$. These equations are all valid for $t \geq 0$.

The equations we just found are generally valid whenever a capacitor C is charged through a resistance R . We see that how quickly we can change the voltage across the capacitor is governed by an RC time constant. If we have a capacitor that is initially charged, and we discharge it through a resistor R , we also get exponential behavior with an RC time constant, although the expressions for $V_C(t)$, $V_R(t)$, $Q(t)$, and $I(t)$ are not the same as the charging case. Working out these expressions for the discharging case will be left as an end-of-chapter problem.

¹More specifically, this is a first-order linear nonhomogeneous differential equation. The general solution to a linear nonhomogeneous differential equation is the sum of the complementary solution (the solution to the homogeneous version of the equation, in this case, the equation with the Q -independent term V_0/R removed) plus the particular solution (a solution that satisfies the original differential equation with no arbitrary constants). In our case, the equation is simple enough that we have opted to skip over this formalism and jump straight to a general form of the solution.

2.2.2 RL CIRCUIT

Next we will consider a circuit consisting of a series combination of a resistor R and an inductor L , known as a series RL circuit, connected via a switch to a DC voltage source V_0 :



As before, we will assume that the switch is closed at time $t = 0$. The resistor voltage is $V_R = I(t)R$ and the inductor voltage is $V_L = L(dI/dt)$. For $t \geq 0$, our loop rule equation is $V_0 - I(t)R - L(dI/dt) = 0$. This can be rearranged as:

$$\frac{dI}{dt} = \frac{V_0}{L} - \frac{R}{L}I(t)$$

The solution for $I(t)$ must give us a constant plus the original function multiplied by a constant, so we use the general exponential solution $I(t) = A + Be^{-t/\tau}$. No current can flow until after the switch is closed, so we must have $I(0) = 0$. Plugging this into our solution gives us $A = -B$, so we can write our solution as $I(t) = A(1 - e^{-t/\tau})$. At $t \rightarrow \infty$ the transient response must die out, leaving us with the DC result. At DC an inductor is just a wire, so our DC result is $I(\infty) = V_0/R$. Plugging this into our solution gives $I(t) = (V_0/R)(1 - e^{-t/\tau})$. We now plug our solution back into the differential equation and solve for τ , which gives us $\tau = L/R$. Our complete solution is then:

$$I(t) = \frac{V_0}{R} (1 - e^{-tR/L})$$

We plug this back into our original expressions for $V_R(t)$ and $V_L(t)$ to find the time-dependent voltages across the resistor and inductor, respectively:

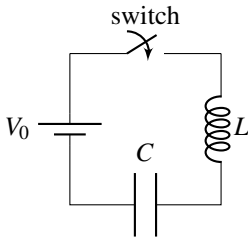
$$V_R(t) = V_0 (1 - e^{-tR/L})$$

$$V_L(t) = V_0 e^{-tR/L}$$

These equations are valid for $t \geq 0$. We see that we cannot change the current through the inductor arbitrarily fast; how quickly the current can change is determined by an L/R exponential time constant.

2.2.3 LC CIRCUIT

Now we will consider a circuit consisting of a series combination of an inductor L and a capacitor C , known as a series LC circuit, connected via a switch to a DC voltage source V_0 :



Again, we assume that the capacitor is initially uncharged and the switch is closed at time $t = 0$. After the switch closes ($t \geq 0$), a current $I(t)$ can flow. We start by writing a loop rule equation for $t \geq 0$: $V_0 - Q(t)/C - L(dI/dt) = 0$. Using $I = dQ/dt$ and rearranging, we can write our loop equation as:

$$\frac{d^2Q}{dt^2} = \frac{V_0}{L} - \frac{Q(t)}{LC}$$

This is a second-order differential equation; our solution $Q(t)$ must be a function that, when we take its second derivative, gives us back a constant plus the original function multiplied by a constant. This solution could be in the form of a sine function, a cosine function, or an exponential function. We will choose a general solution of the form $Q(t) = A + B \sin \omega_0 t + D \cos \omega_0 t$ where A , B , D , and ω_0 are constants to be determined. We know that we must have $I(0) = 0$. Since $I(t) = dQ/dt = \omega_0 B \cos \omega_0 t - \omega_0 D \sin \omega_0 t$, this means that $B = 0$. We must also have $Q(0) = 0$, which means that $A = -D$. Using this, we can simplify our solution to $Q(t) = A(1 - \cos \omega_0 t)$. Now we will plug this solution along with its second derivative back into our original differential equation. When we do this and then bring all the time-dependent terms to the left-hand side of the equation and all the time-independent terms to the right-hand side of the equation, we get $(\omega_0^2 - 1/(LC))A \cos \omega_0 t = (V_0/L - A/LC)$. The only way for the time-dependent left side to be equal to the time-independent right side at all time $t \geq 0$ is for $(\omega_0^2 - 1/(LC)) = 0$, i.e. for $\omega_0 = 1/\sqrt{LC}$. This means the right-hand side must also equal zero, which gives us $A = V_0 C$. Now we can write our final expression for the time-dependent charge:

$$Q(t) = V_0 C (1 - \cos \omega_0 t)$$

with $\omega_0 = 1/\sqrt{LC}$. ω_0 is the angular frequency our system naturally oscillates at when perturbed, and we call it the *resonant frequency*. Expressing it as a linear frequency f_0 , we have $f_0 = \omega_0/(2\pi) = 1/(2\pi\sqrt{LC})$. Using our equation for $Q(t)$, we can then find $V_C(t)$, $V_L(t)$, and $I(t)$:

$$V_C(t) = V_0 (1 - \cos \omega_0 t)$$

$$V_L(t) = V_0 \cos \omega_0 t$$

$$I(t) = V_0 C \omega_0 \sin \omega_0 t$$

Note that $V_L(t)$ and $V_C(t)$ oscillate at the same frequency ω_0 but are π radians out-of-phase. We can see that our circuit obeys the loop rule, $V_L(t) + V_C(t) = V_0$, for all time $t \geq 0$.

Our solution tells us that, after closing the switch, the currents and voltages in the circuit will oscillate forever. This is the electrical equivalent of a perpetual motion machine, i.e. it is impossible. The reason we got this result is that we assumed that all of our circuit components are ideal and have no resistance. In a real circuit, there is always some resistance (from the wire that comprises the inductor, from the voltage source, *etc.*). To model this, we can add a series resistor R that represents the total resistance of the circuit; this makes the circuit a series RLC circuit. The resistance causes the oscillation amplitude to decrease over time, eventually going to zero. Applying the loop rule to the series RLC circuit leads to the differential equation:

$$\frac{d^2 Q}{dt^2} = \frac{V_0}{L} - \frac{R}{L} \frac{dQ}{dt} - \frac{1}{LC} Q(t)$$

Solving this is nontrivial, but in the underdamped case ($1/(LC) > (R/(2L))^2$) it has the solution

$$Q(t) = V_0 C \left[1 - e^{-t/\tau} \left(\cos(\omega t) + \frac{1}{\omega \tau} \sin(\omega t) \right) \right]$$

with $\tau = 2L/R$ and $\omega = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}$. This expression can be used to find the other time-dependent quantities $V_C(t)$, $I(t)$, $V_R(t)$, and $V_L(t)$; working these out will be left as an exercise for the reader. We will revisit the RLC circuit in [section 2.6](#).

2.3 IMPEDANCE

At zero frequency (DC), the ratio of the voltage to the current is called the resistance. At finite frequency (AC), we call the ratio of the time-dependent voltage $V(t)$ to the time-dependent current $I(t)$ the *impedance* Z :

$$Z = \frac{V(t)}{I(t)} \quad (2.11)$$

Like resistance, impedance has units of ohms (Ω). Unlike resistance, impedance is in general a complex number. This reflects the fact that an oscillatory $V(t)$ and $I(t)$ can have both different amplitudes and different phases. A complex number contains two pieces of information – a real part and an imaginary part, or, equivalently, a magnitude and a phase angle – which are used to keep track of the relative amplitudes and the relative phases of the voltage and current.

For oscillatory voltages and currents, we can write $V(t) = V_0 e^{i(\omega t + \phi_V)}$ and $I(t) = I_0 e^{i(\omega t + \phi_I)}$. We assume they have the same frequency ω because the oscillations of voltage and current in a circuit are coupled. Plugging these into Equation 2.11, our impedance is:

$$Z = (V_0/I_0) e^{i(\phi_V - \phi_I)} \quad (2.12)$$

We can express this as a vector on the real-imaginary plane, where the magnitude of the vector $|Z|$ is the ratio of the amplitudes:

$$|Z| = V_0/I_0 \quad (2.13)$$

and the angle of the vector relative to the real axis, ϕ , is the phase difference between the voltage and current:

$$\phi = \phi_V - \phi_I \quad (2.14)$$

which is also known as the phase angle.

Like any complex number, we can write the impedance in the form

$$Z = X + iY \quad (2.15)$$

where X is the real part, Y is the imaginary part, and $i = \sqrt{-1}$ is the imaginary number. (In electrical engineering, the symbol j is often used to represent the imaginary number instead of i . For a brief review of complex numbers, please refer to [Appendix B](#).) X is sometimes called the resistance and Y is called the reactance; both have units of ohms. X and Y are by definition purely real numbers; the imaginary term iY assumes that the i has been factored out, leaving a purely real Y value. When we write the impedance in this form, it makes it easy to find the magnitude

$$|Z| = \sqrt{X^2 + Y^2} \quad (2.16)$$

and the phase angle

$$\phi = \tan^{-1}(Y/X) \quad (2.17)$$

where ϕ is defined from $-\pi/2 \leq \phi \leq \pi/2$.

The impedance of an ideal resistor Z_R is simply its resistance R :

$$Z_R = R \quad (2.18)$$

Since R is always a positive real number, we know that the voltage and current in an ideal resistor must be in-phase, i.e. $\phi_V = \phi_I$.

For a capacitor C with a time-dependent voltage $V(t)$ and a corresponding time-dependent charge $Q(t)$, we can write Equation 2.1 as $V(t) = Q(t)/C$. Taking the time derivative of both sides, we get $dV/dt = (1/C)dQ/dt = (1/C)I(t)$. We can multiply both sides by dt to get $dV = (1/C)I(t)dt$, and then integrating both sides gives $V(t) = (1/C) \int I(t)dt$. If we plug in our expression for the oscillatory current $I(t) = I_0 e^{i(\omega t + \phi_I)}$ and integrate the right-hand side, we get $V(t) = I(t)/(i\omega C)$. Solving for $V(t)/I(t)$, which is the impedance, we get the following expression for the impedance of our capacitor Z_C :

$$Z_C = \frac{1}{i\omega C} \quad (2.19)$$

For an inductor L with a time-dependent current $I(t)$ and a corresponding time-dependent voltage $V(t)$, Equation 2.6 gives us $V(t) = L(dI/dt)$. Plugging in the

derivative of our oscillatory expression for the current $dI/dt = I_0 i \omega e^{i(\omega t + \phi_I)} = i \omega I(t)$, we can write this as $V(t) = i \omega L I(t)$. Hence the impedance of our inductor Z_L is:

$$Z_L = i \omega L \quad (2.20)$$

We note that Z_L is a positive and purely imaginary number, which means that an inductor must have a phase angle $\phi = \tan^{-1}(\infty) = \pi/2$. $Z_C = 1/(i \omega C) = -i/(\omega C)$ is a negative and purely imaginary number, which means that a capacitor must have a phase angle $\phi = \tan^{-1}(-\infty) = -\pi/2$.

Series and parallel impedances combine in the same way that series and parallel resistors combine, except, of course, now we are dealing with complex numbers. If we have N series impedances, the total equivalent impedance is given by:

$$Z_{eq} = \sum_{n=1}^N Z_n \quad (2.21)$$

And if we have N parallel impedances, the total equivalent impedance is given by:

$$Z_{eq} = \left(\sum_{n=1}^N \frac{1}{Z_n} \right)^{-1} \quad (2.22)$$

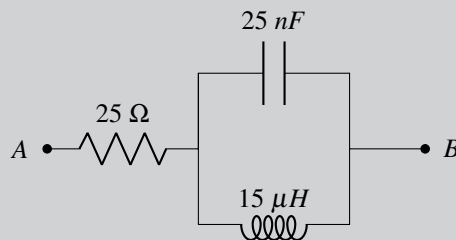
Now that we are assuming oscillatory currents and voltages, we should introduce circuit symbols for oscillatory, i.e. AC, sources. Our symbol for an AC source is:



We will use this symbol for both AC voltage and AC current sources; which one it is should be clear from the context.

Exercise 2.1 Finding the Equivalent Impedance

Find the equivalent impedance of the circuit shown below as measured between points A and B at a frequency $f = 750$ kHz. Express your answer in the form $Z = X + iY$, i.e. separate the real and imaginary parts. Then find the magnitude $|Z|$ and the phase angle ϕ .



Solution: The capacitor and inductor are in parallel, so they have an equivalent impedance:

$$Z_{LC} = \left(\frac{1}{Z_C} + \frac{1}{Z_L} \right)^{-1} = \left(i\omega C + \frac{1}{i\omega L} \right)^{-1} = \frac{-i}{\omega C - \frac{1}{\omega L}}$$

This equivalent impedance is in series with the resistor, so the total equivalent impedance is then:

$$Z_{eq} = Z_R + Z_{LC} = R - \frac{i}{\omega C - \frac{1}{\omega L}}$$

Plugging in numbers, we get $Z_{eq} = 25 - i9.65 \, \Omega$. The real part is $X_{eq} = 25 \, \Omega$ and the imaginary part is $Y_{eq} = -9.65 \, \Omega$. The magnitude is then $|Z_{eq}| = \sqrt{X_{eq}^2 + Y_{eq}^2} = 26.8 \, \Omega$ and the phase angle is $\phi = \tan^{-1}(Y_{eq}/X_{eq}) = -0.368 \, \text{rad} = -21.1^\circ$.

2.4 POWER IN AC CIRCUITS

We can extend our expression for power dissipation in DC circuits (Equation 1.12) to the AC case by making all the quantities time-dependent. However, we must ensure that power is always a purely real number. To do this, we find the time-dependent power $P(t)$ by taking the product of the real part of the time-dependent current $\text{Re}[I(t)]$ and the real part of the time-dependent voltage $\text{Re}[V(t)]$:

$$P(t) = \text{Re}[I(t)] \text{Re}[V(t)] \quad (2.23)$$

We call $P(t)$ the instantaneous power. In AC circuits, we are often interested in finding the time-average power. We denote the time-average power as $P_{avg} = \langle P(t) \rangle$.

If we plug in our oscillatory expressions for the current $I(t) = I_0 e^{i(\omega t + \phi_I)}$ and the voltage $V(t) = V_0 e^{i(\omega t + \phi_V)}$ into Equation 2.23 and apply Euler's formula, we get:

$$P(t) = I_0 V_0 \cos(\omega t + \phi_I) \cos(\omega t + \phi_V)$$

Using the product-to-sum identity, this becomes:

$$P(t) = \frac{1}{2} I_0 V_0 [\cos(2\omega t + \phi_I + \phi_V) + \cos \phi]$$

We see that there is a term that oscillates at an angular frequency 2ω , i.e. twice the frequency of the AC current and voltage, and a DC term that depends on the phase angle $\phi = \phi_V - \phi_I$. If we take the time-average of this expression, the oscillatory component time-averages to zero and we have:

$$P_{avg} = \frac{1}{2} I_0 V_0 \cos \phi$$

I_0 and V_0 are zero-to-peak amplitudes. We introduce the root-mean-square (rms) amplitudes $I_{rms} = I_0/\sqrt{2}$ and $V_{rms} = V_0/\sqrt{2}$, which allows us to write the time-average power as:

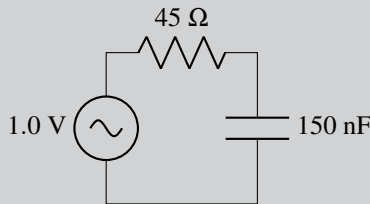
$$P_{avg} = I_{rms} V_{rms} \cos \phi \quad (2.24)$$

When specifying the amplitude of an AC signal, it is common to specify it as an rms amplitude.

For a resistor, $\phi = 0$ and hence $P_{avg} = I_{rms} V_{rms}$. For an ideal inductor, $\phi = \pi/2$ and so $P_{avg} = 0$. Likewise, for an ideal capacitor, $\phi = -\pi/2$ and so $P_{avg} = 0$. We see that only resistors dissipate time-average power; ideal inductors and capacitors do not. Note that the instantaneous power is not zero for an inductor or a capacitor. Half the time, the inductor and capacitor are absorbing energy from the circuit, and the other half of the time, they are emitting that energy back into the circuit.

Exercise 2.2 Finding the Time-Average Power

Find the time-average power dissipated in the circuit shown. The AC voltage source has an rms amplitude of 1.0 V and a frequency $f = 27$ kHz.

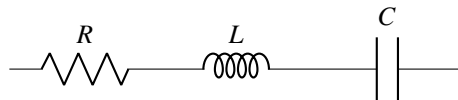


Solution: First we need to find the total impedance of the resistor and capacitor. They are in series, so $Z_{eq} = Z_R + Z_C = R - i/(\omega C) = 45 - i39.3 \Omega$. We can use this to find the corresponding magnitude $|Z_{eq}| = 59.7 \Omega$ and phase angle $\phi = -0.718$ rad. The rms amplitude of the current through the circuit can be found from $I_{rms} = V_{rms}/|Z_{eq}| = 0.0167$ A. Then the time average power can be found using Equation 2.24: $P_{avg} = (0.0167\text{A})(1.0\text{V}) \cos(-0.718) = 12.6$ mW.

2.5 RESONANT CIRCUITS

2.5.1 SERIES RLC CIRCUIT

A series combination of a resistor R , inductor L , and capacitor C is known as a series RLC circuit:



In [section 2.2.3](#), we saw that such a circuit has a characteristic or resonant frequency $f_0 = \omega_0/(2\pi) = 1/(2\pi\sqrt{LC})$. The impedance of the series RLC circuit is

$$Z_{eq} = Z_R + Z_L + Z_C = R + i(\omega L - 1/(\omega C))$$

The magnitude is

$$|Z_{eq}| = \sqrt{R^2 + (\omega L - 1/(\omega C))^2}$$

and the phase angle is

$$\phi = \tan^{-1}((\omega L - 1/(\omega C))/R).$$

Plots of Z_{eq} and ϕ as functions of ω are shown in [Figure 2.1](#). We see that the impedance magnitude reaches a minimum at the resonant frequency $\omega_0 = 1/\sqrt{LC}$, and the value of this minimum is R . At the resonant frequency, we also have $\phi = 0$. This arises because the inductive impedance and capacitive impedance cancel at the resonant frequency, and hence the circuit behaves like just a resistor.

We now consider the time-average power P_{avg} dissipated in our series RLC circuit when it is connected to an AC voltage source with rms amplitude V_{rms} . The rms current amplitude is then $I_{rms} = V_{rms}/|Z_{eq}| = V_{rms}/\sqrt{R^2 + (\omega L - 1/(\omega C))^2}$ and the phase angle is $\phi = \tan^{-1}((\omega L - 1/(\omega C))/R)$. We can then plug these into Equation 2.24 to find P_{avg} . It is helpful to remember the trigonometry identity $\cos(\tan^{-1}x) = 1/\sqrt{1+x^2}$. Using this, we find

$$P_{avg} = \frac{V_{rms}^2 R}{R^2 + (\omega L - 1/(\omega C))^2}$$

A plot of P_{avg} as a function of ω is shown in [Figure 2.2](#). We see that P_{avg} reaches a maximum value of V_{rms}^2/R at the resonant frequency $\omega = \omega_0 = 1/\sqrt{LC}$, and $P_{avg} \rightarrow 0$ in the limits $\omega \rightarrow 0$ and $\omega \rightarrow \infty$.

The quality factor Q of a resonant circuit is defined as

$$Q = \frac{\omega_0}{\Delta\omega} \tag{2.25}$$

where ω_0 is the resonant frequency and $\Delta\omega$ is the full-width at half-maximum (FWHM) of P_{avg} as a function of ω ([Figure 2.2](#)). Q is a unitless measure of how “sharp” a resonance is, with a larger Q corresponding to a sharper resonance. A resonator with $Q > 1/2$ is underdamped, a resonator with $Q < 1/2$ is overdamped, and a resonator with $Q = 1/2$ is critically damped.

We can find an expression for Q for our series RLC circuit by setting the expression for P_{avg} given above equal to $V_{rms}^2/(2R)$, i.e. half its maximum value, and solving for the two values of ω that satisfy this equation. $\Delta\omega$ is then the difference between the larger and the smaller of these two values. When we do this, we get $\Delta\omega = R/L$. Plugging this and our expression for ω_0 into Equation 2.25, we find the quality factor of our series RLC circuit:

$$Q = \frac{1}{R} \sqrt{\frac{L}{C}}$$

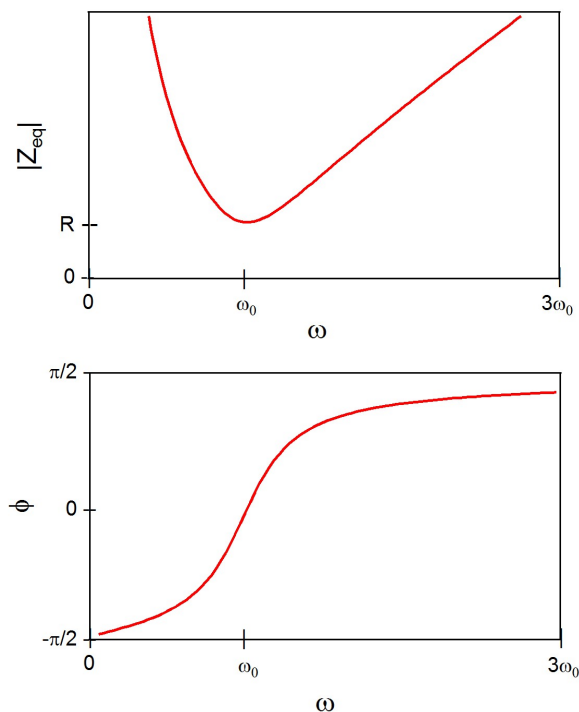
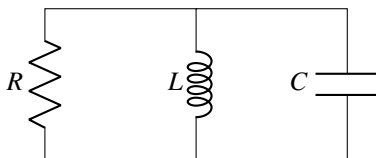


Figure 2.1 The impedance magnitude $|Z_{eq}|$ and phase angle ϕ of a series RLC circuit with a quality factor $Q = 2$ as functions of frequency. The resonant frequency is $\omega_0 = 1/\sqrt{LC}$.

2.5.2 PARALLEL RLC CIRCUIT

A parallel combination of a resistor R , an inductor L , and a capacitor C is known as a parallel RLC circuit:



The impedance of the parallel RLC circuit is:

$$Z_{eq} = \left(\frac{1}{Z_R} + \frac{1}{Z_L} + \frac{1}{Z_C} \right)^{-1} = \frac{1}{1/R + i(\omega C - 1/(\omega L))}$$

To separate the real and imaginary parts of Z_{eq} , we multiply both the numerator and the denominator by the complex conjugate of the denominator. After doing this and

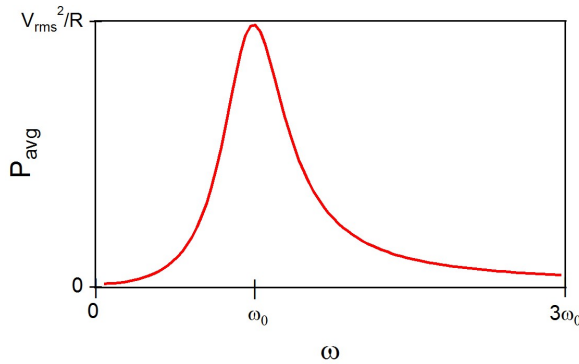


Figure 2.2 The time-average power P_{avg} dissipated in a series RLC circuit with $Q = 2$ connected to an AC voltage source with amplitude V_{rms} as a function of frequency.

simplifying, we get:

$$Z_{eq} = \frac{1/R}{(1/R)^2 + (\omega C - 1/(\omega L))^2} - i \frac{\omega C - 1/(\omega L)}{(1/R)^2 + (\omega C - 1/(\omega L))^2}$$

Solving for the magnitude, we get:

$$|Z_{eq}| = \frac{1}{\sqrt{(1/R)^2 + (\omega C - 1/(\omega L))^2}}$$

and solving for the phase angle, we get:

$$\phi = \tan^{-1} (R(1/(\omega L) - \omega C))$$

Plots of $|Z_{eq}|$ and ϕ as functions of ω are shown in [Figure 2.3](#). We see that $|Z_{eq}|$ is a maximum at the resonant frequency $\omega_0 = 1/\sqrt{LC}$ and that its maximum value is R . $\phi = 0$ at the resonant frequency. Like the series RLC, the parallel RLC behaves like just a resistor on resonance. The difference is that this is a maximum of the impedance for the parallel RLC, while it is a minimum of the impedance for the series RLC.

If we connect the parallel RLC circuit to an AC voltage source with rms amplitude V_{rms} , the time average power is simply $P_{avg} = V_{rms}^2/R$. This is because an ideal voltage source will always maintain the specified voltage across its terminals, regardless of the impedance, and we know that only the resistor will dissipate time-average power.

A more interesting scenario is connecting an AC current source with rms amplitude I_{rms} to our parallel RLC. In this case, we can find the rms voltage across our circuit from $V_{rms} = I_{rms}|Z_{eq}|$, and then $P_{avg} = I_{rms}V_{rms} \cos \phi$, or, recognizing that parallel components have the same voltage and only the resistor dissipates time-average power, we can also write $P_{avg} = V_{rms}^2/R$. Plugging in our expressions from above and simplifying, we get:

$$P_{avg} = \frac{I_{rms}^2 R}{1 + R^2(\omega C - 1/(\omega L))^2}$$

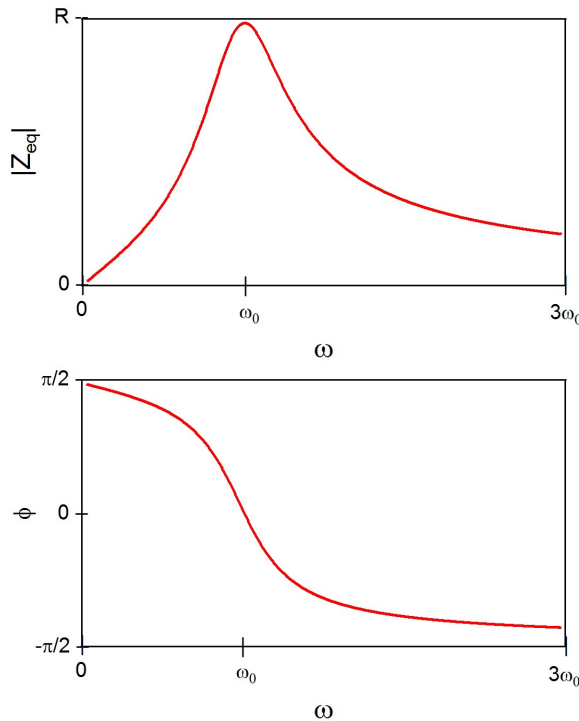


Figure 2.3 The impedance magnitude $|Z_{eq}|$ and phase angle ϕ of a parallel RLC circuit with $Q = 2$ as functions of frequency.

A plot of P_{avg} as a function of ω is shown in [Figure 2.4](#). We see that, like the series RLC case, P_{avg} is a maximum at the resonant frequency, with a maximum value of $I_{rms}^2 R$. The quality factor can be found using the same approach as we used for the series RLC case, which yields for the parallel RLC case:

$$Q = R\sqrt{\frac{C}{L}}$$

2.6 THE OSCILLOSCOPE

Oscilloscopes measure a time-dependent voltage for some specified time interval. Key to using an oscilloscope is the concept of triggering. The trigger is a user-specified voltage, and the time at which the measured signal first crosses the trigger voltage is defined as time $t = 0$ for each measurement. This zero is placed by default on the center of the time axis, but on most oscilloscopes, it is possible to offset it to the left or the right. It is generally possible to trigger off the signal being measured or a separate signal, which is known as external triggering. For a multi-channel oscilloscope, which measures multiple voltages with a common time axis, one can trigger off of any one of the measured channels.

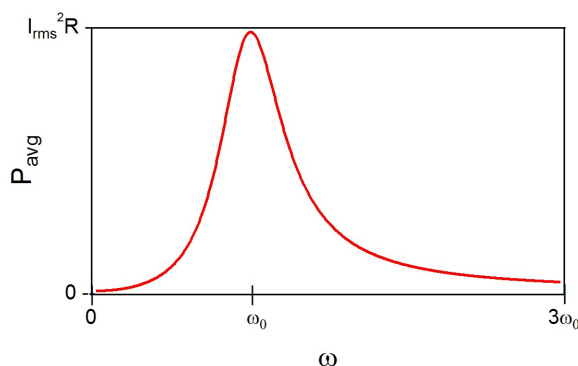


Figure 2.4 The time-average power P_{avg} dissipated in a parallel RLC circuit with $Q = 2$ connected to an AC current source with amplitude I_{rms} as a function of frequency.

Most oscilloscopes can be set to trigger off a rising edge – meaning that the instrument looks for the first time in the measurement window when the signal goes from below to above the trigger level – or to trigger off a falling edge – meaning that the instrument looks for the first time in the measurement window that the signal goes from above to below the trigger level. If you are measuring a periodic waveform and the waveform appears to be wandering back and forth along the time axis, this is a sign that the oscilloscope isn't triggering properly, which results in an arbitrary definition of $t = 0$ for each measurement iteration. There are two standard triggering modes: normal and auto. In normal mode, the oscilloscope display only updates if the trigger threshold is crossed. In auto mode, the oscilloscope display updates on each measurement iteration regardless of whether or not the trigger threshold is crossed; if the trigger threshold is not crossed, the instrument chooses an arbitrary definition of $t = 0$.

A representative example of an oscilloscope display is shown in [Figure 2.5](#). This is from a two-channel oscilloscope. Channel 1 shows a square wave signal and channel 2 shows a sinusoidal signal. The oscilloscope was set to trigger on channel 2. The trigger level is indicated by the arrow on the right edge of the display. The trigger settings are shown in the bottom right of the display; there one can see that it is a rising edge trigger set to a level of 0 V on channel 2. The measured frequency of the signal being used to trigger is also shown, which in this case happens to be 1.134 kHz. Along the bottom of the display are shown the channel 1 and channel 2 voltage scales, where the value specifies the number of volts per division, with the divisions indicated by the gray dashed grid lines. The time scale is also shown, expressed as time per division. The voltage scales can be set to different values for the two channels, but the two channels share the same time scale. The ground (0 V) position for each channel is indicated by the arrows on the left side of the display. Here, the grounds have been offset so the waveforms on the two channels do not overlap.

The channel 1 and channel 2 inputs to the oscilloscope are usually coaxial BNC connectors.² The outer conductors, or grounds, of these BNC connectors are electrically connected inside the instrument and are also connected to line ground via the third prong in the instrument’s power cable. This means that, when using more than one channel, the grounds of both channels must be connected to points in the circuit that are the same electric potential. If they are not, then the grounds will create a short-circuit inside the oscilloscope between the two different points at which they are connected.

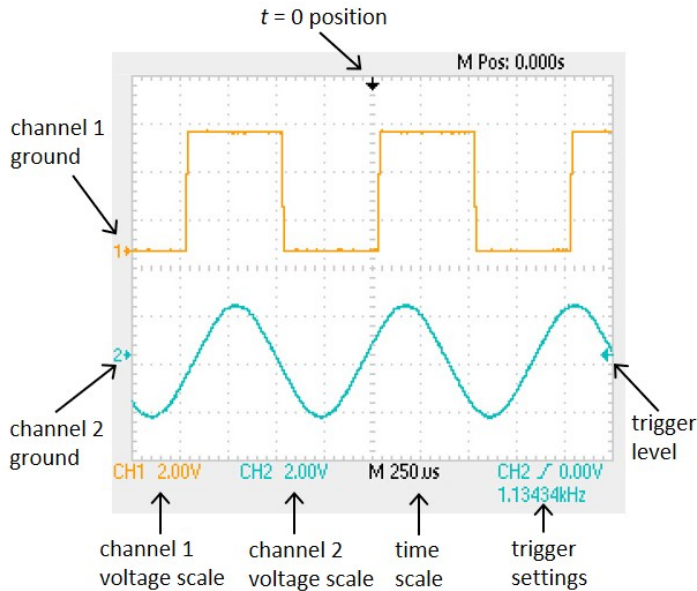


Figure 2.5 Display of a two-channel oscilloscope.

It is common to use an oscilloscope probe to connect the oscilloscope to the circuit being measured. An oscilloscope probe usually has a BNC connector on one end to connect to the oscilloscope, and the other end typically has a retractable clip and a mini alligator clip. The retractable clip connects to the center conductor of the BNC connector and the mini alligator clip connects to the outer conductor or ground of the BNC connector. Oscilloscope probes are not just a pair of wires; they have a voltage divider built into the probe near the point at which it connects to the circuit. Some probes have a fixed divide-by-ten (10X) voltage divider and some have a voltage divider that can be set to different values such as 1X, 10X, and 100X. These divide-by factors are sometimes referred to as probe attenuation.

²BNC is an acronym for Bayonet Neill Concelman, where bayonet refers to the locking mechanism and Neill and Concelman refer to its two inventors. BNC connectors work well up to frequencies of approximately 1 – 2 GHz, while the use of higher frequencies generally necessitates a different type of connector.

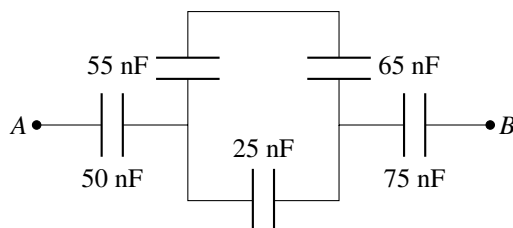
The reason for the voltage divider is to minimize the effect of the capacitance of the probe and the oscilloscope on the measurement. Ideally, the oscilloscope would behave as an ideal voltmeter, meaning that it would have infinite input impedance. In practice, the oscilloscope has a large but finite input resistance, often $10\text{ M}\Omega$. It also has some input capacitance, typically around 20 pF . The probes also have some capacitance between their two conductors, $\approx 20\text{ pF}$ per foot of probe length. The total capacitance is thus usually in the range of $50 - 100\text{ pF}$. Without a voltage divider, the impedance at the point of connection between the probe and the circuit is then the parallel combination of a $R = 10\text{ M}\Omega$ resistor and a $C = 50 - 100\text{ pF}$ capacitor. The magnitude of this impedance is $|Z| = ((1/R^2) + (\omega C)^2)^{-1/2}$. As the measurement frequency increases, the capacitance causes the input impedance $|Z|$ to decrease. If $|Z|$ is not much larger than the equivalent impedance of what is being measured, then the oscilloscope no longer behaves as a quasi-ideal voltmeter. This means that connecting the oscilloscope probe to the circuit will alter the behavior of the circuit.

The voltage divider in the oscilloscope probe, located close to where the probe connects to the circuit, presents a large resistive impedance to the circuit along with minimal capacitance. The output of the voltage divider has a small resistive impedance, helping to ensure that the oscilloscope remains in the quasi-ideal limit even as the capacitance lowers the oscilloscope input impedance at higher frequencies. It is important that the oscilloscope knows what voltage divider the probe is using, as it will compensate for the voltage divider in its display. For example, if a $10X$ probe is used, the oscilloscope display will scale up the measured voltage by a factor of 10 to compensate for the factor of 10 lost in the probe. When using oscilloscope probes, one should always make sure that each channel is set to an attenuation factor that is equal to the divide-by factor of the probe being used for that channel. If one is using an ordinary cable and not an oscilloscope probe, then there is no voltage divider and one should make sure the attenuation factor for that oscilloscope channel is set to $1X$.

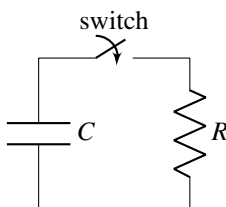
Oscilloscopes that are designed to operate at microwave frequencies ($\sim\text{GHz}$) often have a $50\text{ }\Omega$ input impedance rather than a $\approx 10\text{ M}\Omega$ input impedance. The reasons for this choice of input impedance are discussed in [Chapter 4](#).

2.7 PROBLEMS

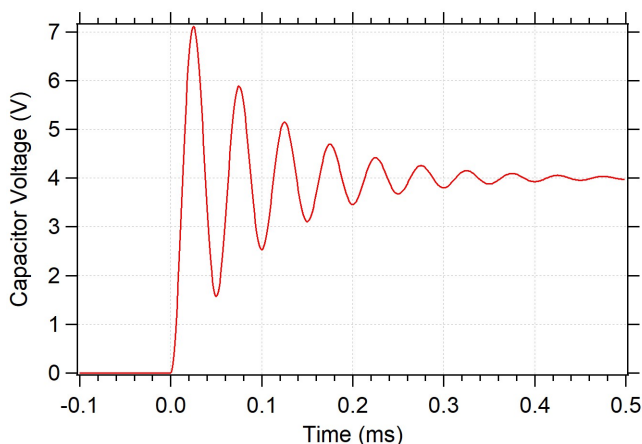
Problem 1. (a) Find the total equivalent capacitance between points A and B . (b) If points A and B are connected to opposite sides of a 1.5 V battery, find the voltage across each individual capacitor.



Problem 2. In the RC circuit shown below, the capacitor initially has a voltage V_0 and the switch is initially open. At time $t = 0$, the switch is closed, allowing the capacitor to discharge through the resistor. (a) Use the loop rule to write a differential equation valid for all time $t \geq 0$ in which the capacitor charge $Q(t)$ is the only time-dependent quantity. (Hint: dQ/dt is negative for the discharging case, so we should write the current as $I = -dQ/dt$.) (b) Solve this differential equation to get an expression for $Q(t)$ valid for $t \geq 0$ that depends only on the time t , the capacitance C , the resistance R , and the initial capacitor voltage V_0 . (c) Find the corresponding expressions for the time-dependent capacitor voltage $V_C(t)$, resistor voltage $V_R(t)$, and resistor current $I(t)$.



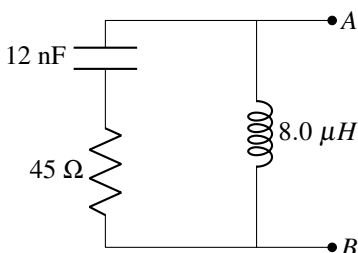
Problem 3. The plot below shows the voltage across a capacitor that is part of a series RLC circuit with an inductance $L = 6.3 \mu\text{H}$. The capacitor is initially uncharged, and at time $t = 0$, the series RLC circuit is connected to a battery. Assuming that we are in the underdamped regime ($1/(LC) > (R/(2L))^2$), use the plot to estimate (a) the battery voltage, (b) the capacitance C , and (c) the resistance R .



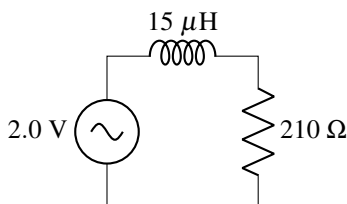
Problem 4. An 85Ω resistor is in parallel with a 12 nF capacitor. What is the equivalent impedance of these two parallel components at a frequency $f = 150 \text{ kHz}$? Express your answer in the form $Z = X + iY$, i.e. separate the real and imaginary parts.

Problem 5. You have a 27 nH inductor. What value of capacitor should you put in series with the inductor in order to achieve an impedance of $Z = i60\ \Omega$ at $f = 1.2\text{ GHz}$?

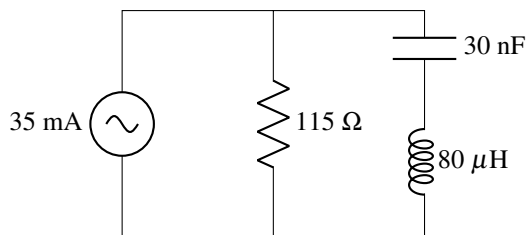
Problem 6. Find the impedance Z between points A and B at a frequency $f = 1.4\text{ MHz}$. Be sure to separate the real and imaginary parts. Then find the impedance magnitude $|Z_{eq}|$ and the phase angle ϕ .



Problem 7. In the circuit shown, the AC voltage source has an rms amplitude of 2.0 V and a frequency $f = 5.2\text{ MHz}$. What is the time-average power dissipated in the circuit?



Problem 8. In the circuit shown, the AC current source has an rms amplitude of 35 mA and a frequency $f = 1.2\text{ MHz}$. What is the time-average power dissipated in the circuit?





Taylor & Francis

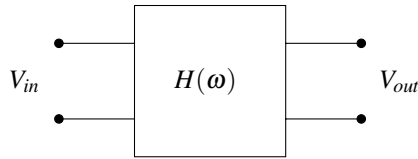
Taylor & Francis Group

<http://taylorandfrancis.com>

3 Four-Terminal Circuits and the Transfer Function

3.1 THE TRANSFER FUNCTION

A four-terminal circuit is one in which an input voltage V_{in} is applied between two input terminals and this produces an output voltage V_{out} between two output terminals. A general representation of a four-terminal circuit is shown below, with the input terminals on the left and the output terminals on the right:



The box in the center is a placeholder for some circuit that connects the input terminals to the output terminals. We define the ratio of the output voltage to the input voltage as the transfer function H :

$$H(\omega) = \frac{V_{out}}{V_{in}} \quad (3.1)$$

Like impedance, the transfer function is in general frequency-dependent and complex. This reflects the fact that it contains two pieces of information: the relative amplitudes of V_{in} and V_{out} and their relative phases. Just like for impedance, we can express H in the form of a real part H_R and an imaginary part H_I :

$$H = H_R + iH_I \quad (3.2)$$

The magnitude of the transfer function $|H|$ is equal to the ratio of the amplitude of V_{out} to the amplitude of V_{in} and can be found from:

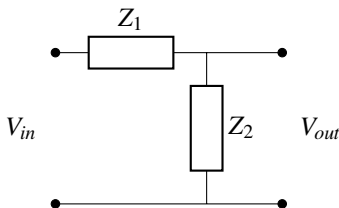
$$|H| = \sqrt{H_R^2 + H_I^2} \quad (3.3)$$

The phase angle of the transfer function ϕ_H is equal to the phase difference between the output voltage and the input voltage and can be found from:

$$\phi_H = \tan^{-1}(H_I/H_R) \quad (3.4)$$

3.2 SIMPLE FILTER CIRCUITS

One example of a four-terminal circuit that we've already encountered is the voltage divider. We can rewrite Equation 1.10 to get the transfer function for our DC voltage divider: $H = R_2/(R_1 + R_2)$. We can generalize the voltage divider by replacing resistances R_1 and R_2 with impedances Z_1 and Z_2 :

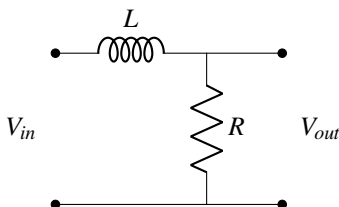


The transfer function for our voltage divider is then:

$$H(\omega) = \frac{Z_2}{Z_1 + Z_2} \quad (3.5)$$

Z_1 and Z_2 may be the impedance of an individual component or they may be the equivalent impedance of multiple components.

As an example, let's consider the following circuit:



Using the inductor's impedance $i\omega L = Z_1$ and the resistor's impedance $R = Z_2$ in Equation 3.5, we get:

$$H = \frac{R}{R + i\omega L}$$

To separate the real and imaginary parts, we multiply both the numerator and the denominator by the complex conjugate of the denominator. When we do this and simplify, we get:

$$H = \frac{R^2}{R^2 + (\omega L)^2} - i \frac{\omega RL}{R^2 + (\omega L)^2}$$

We can then find the magnitude and phase angle:

$$|H| = \frac{1}{\sqrt{1 + (\omega L/R)^2}}$$

$$\phi_H = \tan^{-1} \left(\frac{-\omega L}{R} \right)$$

In the DC limit ($\omega \rightarrow 0$), we get $|H| = 1$ and $\phi_H = 0$. In the high-frequency limit ($\omega \rightarrow \infty$), we get $|H| = 0$ and $\phi_H = -\pi/2$. In [Figure 3.1](#), we plot $|H|$ and ϕ_H as functions of ω . We see that our circuit passes low-frequency signals and does not pass high-frequency signals. We call a circuit with this type of behavior a *low-pass filter*. More specifically, our circuit is an RL low-pass filter.

$|H(\omega)|$ is sometimes plotted on a logarithmic scale, where our y-axis becomes $20\log|H|$ and we indicate this log scaling by using the label dB, which stands for decibels. This is shown in Figure 3.2. The plots of Figures 3.1 and 3.2, which show the frequency response of our system, are known as Bode plots.

It is common to express the bandwidth of our low-pass filter using the frequency at which $|H|$ decreases by a factor of $1/\sqrt{2}$ or, on a log scale, -3 dB. This frequency is sometimes referred to as the “3 dB point.” For our LC low-pass filter, the 3 dB point occurs at a frequency $\omega = R/L$. At this frequency, the phase angle is $\phi_H = -\pi/4$.

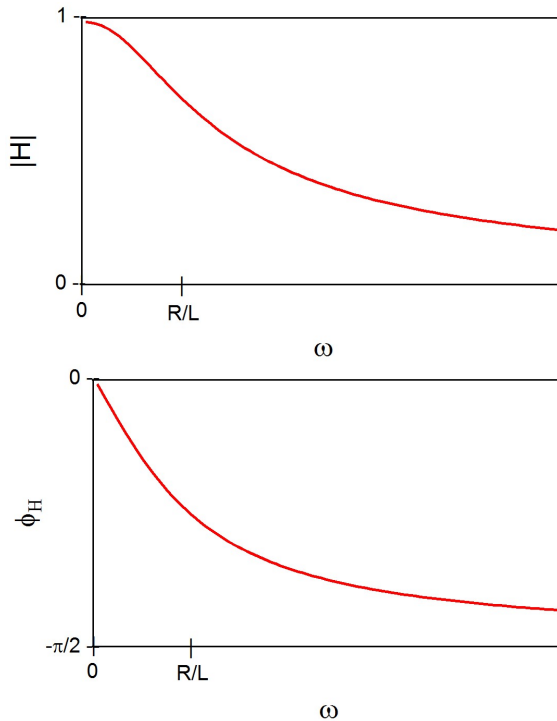


Figure 3.1 Magnitude of the transfer function $|H|$ and phase angle ϕ_H for our RL low-pass filter as functions of frequency.

The region of frequency where $|H| \approx 1$ is known as the filter pass band and the region of frequency where $|H| \ll 1$ is known as the filter stop band. Our LC low-pass filter is a single-stage filter because it is composed of a single series impedance (the inductor) and a single impedance to ground (the resistor). A multi-stage filter is comprised of a sequence of multiple series impedances alternating with impedances to ground. Multi-stage filters are capable of achieving a greater slope for $|H(\omega)|$, i.e. a more abrupt transition between the pass band and the stop band. However, the analysis of multi-stage filters is more complex. Here, for the sake of simplicity, we will restrict ourselves to considering single-stage filters that can be modeled as

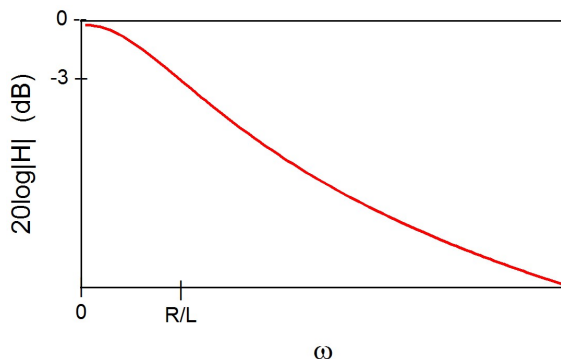


Figure 3.2 Plot of $20 \log |H|$ as a function of frequency for our RL low-pass filter.

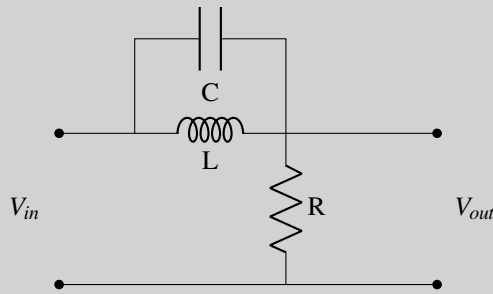
voltage dividers using Equation 3.5.

There are four general types of filters, of which the low-pass is one. The other three are the high-pass, the band-pass, and the band-stop. As the names suggest, a high-pass filter passes high frequencies and does not pass low frequencies, a band-pass filter passes some range of frequencies but does not pass frequencies above or below this range, and a band-stop filter does not pass some range of frequencies but passes frequencies above and below this range. We can usually tell which type of filter we have by evaluating $|H|$ in the zero frequency and infinite frequency limits.

Real filters are not as perfect as models based on ideal components would suggest. For example, our RL low-pass filter model predicts $|H| \approx 0$ for $\omega \gg R/L$, regardless of how high we go in frequency. A real low-pass filter always has a finite stop band, and if we go too high in frequency, we will start to see $|H|$ deviate from our model behavior. This arises because the impedance of our components is more complicated than our ideal assumptions. For example, a solenoid inductor has some capacitance between its windings. The contribution of this capacitance to the inductor's impedance becomes significant if we go too high in frequency. Real inductors often specify a self-resonance frequency, which is the frequency at which the inductor's capacitance resonates with its inductance, $\omega_0 = 1/\sqrt{LC}$. Often one can treat an inductor as quasi-ideal if one stays well below this self-resonance frequency. When considering filters in real applications, it is a good idea to keep their nonideality in mind.

Exercise 3.1 Filter Analysis

Find the transfer function and its magnitude and phase angle for the filter circuit shown. Then specify what type of filter this is.



Solution: In our AC voltage divider formula (Equation 3.5), we have $Z_2 = R$ and Z_1 is the equivalent impedance of the inductor and capacitor in parallel, $Z_1 = [i(\omega C - 1/(\omega L))]^{-1}$. So we have:

$$\begin{aligned} H &= \frac{Z_2}{Z_2 + Z_1} = \frac{R}{R - \frac{i}{\omega C - 1/(\omega L)}} \\ &= \frac{R^2}{R^2 + (\omega C - 1/(\omega L))^{-2}} - i \frac{R/(\omega C - 1/(\omega L))}{R^2 + (\omega C - 1/(\omega L))^{-2}} \end{aligned}$$

For the magnitude, we get:

$$|H| = \frac{R}{\sqrt{R^2 + 1/(\omega C - 1/(\omega L))^2}}$$

And for the phase angle we get:

$$\phi_H = \tan^{-1} \left(\frac{1}{R(\omega C - 1/(\omega L))} \right)$$

In the limit $\omega \rightarrow 0$, we get $|H| \rightarrow 1$ and $\phi_H \rightarrow 0$. In the limit $\omega \rightarrow \infty$, we get $|H| \rightarrow 1$ and $\phi_H \rightarrow 0$. Thus it would appear that we have a band-stop filter. To check this, we can evaluate $|H|$ and ϕ_H at the resonant frequency $\omega_0 = 1/\sqrt{LC}$, for which we get $|H| = 0$ and $\phi_H = \pm\pi/2$. Approaching the resonant frequency from below, $\phi_H \rightarrow -\pi/2$, while approaching the resonant frequency from above, $\phi_H \rightarrow \pi/2$. The stop band of our band-stop filter is centered at the resonant frequency, and the width of the stop band is set by the quality factor of the parallel LC resonator.

3.3 FOURIER ANALYSIS

Both the impedance and the transfer function were defined in terms of time-dependent voltages and/or currents, yet in both cases, we naturally arrive at frequency-dependent expressions when we assume oscillatory time-dependent

signals. In working with these expressions, it is useful to be able to transform a signal from a time-domain representation to a frequency-domain representation and vice versa. This can be done using the Fourier transform. Before we describe this, we will first discuss a related mathematical concept known as a Fourier series.

3.3.1 FOURIER SERIES

Any periodic function $f(t)$ can be expressed as a sum of single-frequency sine and cosine functions with different weights. This sum is known as a Fourier series. If the period of $f(t)$ is T , then the Fourier series for $f(t)$ is given by

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \right] \quad (3.6)$$

where weighting coefficients a_n and b_n are given by

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos\left(\frac{2\pi nt}{T}\right) dt \quad \text{with } n = 0, 1, 2, \dots \quad (3.7)$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin\left(\frac{2\pi nt}{T}\right) dt \quad \text{with } n = 1, 2, 3, \dots \quad (3.8)$$

The frequency of each sine and cosine term is $\omega = 2\pi n/T$, so each finite frequency term is an integer multiple of the fundamental frequency $\omega_f = 2\pi/T$. The DC component $a_0/2$ is equal to the time-average of the original function $f(t)$.

As an example of finding the Fourier series for a function, we'll consider the square wave function $f(t)$ shown in [Figure 3.3](#). $f(t)$ is an odd function because $f(-t) = -f(t)$. It has a zero-to-peak amplitude of 1 and a period $T = 1$.

The time-average of our function is zero, hence $a_0 = 0$. Additionally, *cosine* is an even function, the product of an even function and an odd function is an odd function, and the integral of an odd function over a period is zero, so $a_n = 0$. This leaves only the b_n coefficients to evaluate. We can do this by dividing our integral into two regions, from $-T/2$ to 0 where $f(t) = -1$ and from 0 to $T/2$ where $f(t) = 1$:

$$b_n = -\frac{2}{T} \int_{-T/2}^0 \sin \frac{2n\pi t}{T} dt + \frac{2}{T} \int_0^{T/2} \sin \frac{2n\pi t}{T} dt = \frac{2}{n\pi} [1 - \cos(n\pi)]$$

For n even, $\cos(n\pi) = 1$ and hence $b_n = 0$, and for n odd, $\cos(n\pi) = -1$, and hence we have:

$$b_n = \frac{4}{n\pi} \quad \text{with } n = 1, 3, 5, \dots$$

Redefining n as $(2n+1)$ so we can take the sum over $n = 0, 1, 2, 3, \dots$, we can then write our Fourier series for $f(t)$ as:

$$f(t) = \sum_{n=0}^{\infty} \frac{4}{(2n+1)\pi} \sin\left(\frac{2(2n+1)\pi t}{T}\right)$$

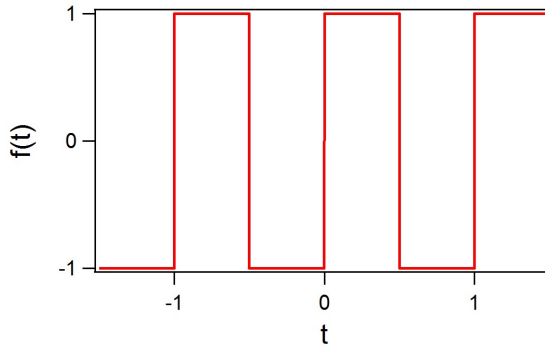


Figure 3.3 Square wave function $f(t)$ with a period $T = 1$.

We see that a square wave of frequency $f_0 = 1/T$ contains frequency components at $f_0, 3f_0, 5f_0, \dots$. As an example, we plot in [Figure 3.4](#) the original square-wave function along with the Fourier series up to $n = 2$ and $n = 10$ for comparison.

The Fourier series can be useful for understanding the effect of passing a periodic signal through a circuit with some transfer function. For example, consider passing the square wave voltage shown in [Figure 3.3](#) through a low-pass filter with a transfer function $H(f) = e^{-2\pi f/100}$. In our Fourier series, we identify the frequency of each Fourier component as $f = (2n+1)/T$, and hence we can write our transfer function in terms of n as $H(f) = \sum_{n=0}^{\infty} e^{-2\pi(2n+1)/(100T)}$. The signal at the filter output $V_{out}(t)$ is then the original Fourier series multiplied by this transfer function, which gives us

$$V_{out}(t) = \sum_{n=0}^{\infty} \frac{4}{(2n+1)\pi} \sin\left(\frac{2(2n+1)\pi t}{T}\right) e^{-\frac{2\pi(2n+1)}{100T}}$$

In [Figure 3.5](#), we plot $V_{out}(t)$ with $T = 1$ along with the original square wave signal for comparison.¹

3.3.2 FOURIER TRANSFORM

The Fourier transform allows us to convert a time-domain function into a frequency-domain function and vice versa, even if the function is nonperiodic. If we have a time-domain function $f(t)$, its frequency-domain representation $f(\omega)$ is given by the Fourier transform:

$$f(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (3.9)$$

¹A question for the reader: Why does the low-pass filter smooth out our square wave ([Figure 3.5](#)) rather than producing the ringing response that we see when we truncate the Fourier series for our square wave ([Figure 3.4](#))?

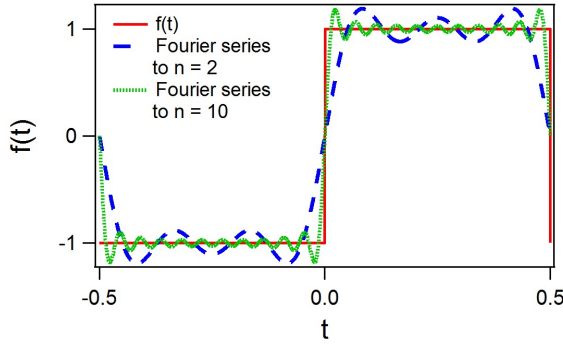


Figure 3.4 Square wave function $f(t)$ with period $T = 1$ along with its Fourier series up to $n = 2$ and $n = 10$.

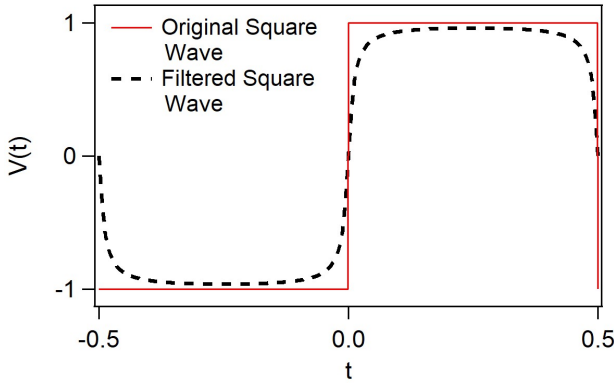


Figure 3.5 Square wave $V(t)$ with period $T = 1$ before and after passing through a low-pass filter.

Note that the units of $f(\omega)$ are the units of $f(t)$ multiplied by a unit of time. The reverse Fourier transform, which allows us to convert a frequency-domain function $f(\omega)$ to a time-domain function $f(t)$, is:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega) e^{i\omega t} d\omega \quad (3.10)$$

Sometimes, instead of having the factor of $1/(2\pi)$ in the reverse transform as we have done, a factor of $1/\sqrt{2\pi}$ is used in both the transform and its reverse. Most mathematical and data analysis software packages have a built-in Fourier transfer function that is based on the numerical procedure known as a fast Fourier transform (FFT). This numerical procedure is usually much more computationally efficient than evaluating the integrals in Equations 3.9 and 3.10.

One can use the Fourier transform to determine how a time-domain signal will be changed by passing through a circuit with some specified transfer function. To do this, one takes the Fourier transform of the input signal $V_{in}(t)$ to get $V_{in}(\omega)$, multiplies this by the transfer function $H(\omega)$ to get $V_{out}(\omega)$, and then takes the reverse Fourier transform to get $V_{out}(t)$. If we only care about the amplitudes of $V_{in}(t)$ and $V_{out}(t)$ and not their relative phase, then we can just multiply by the magnitude of the transfer function $|H(\omega)|$, which generally makes the task simpler.

As an example, we will consider the effect of passing a Gaussian voltage pulse through an RL low-pass filter. Our initial voltage pulse will be $V_{in}(t) = V_0 e^{-at^2}$. The Fourier transform of this is:

$$V_{in}(\omega) = V_0 \int_{-\infty}^{\infty} e^{-at^2} e^{-i\omega t} dt$$

We can use Euler's formula to write this as:

$$V_{in}(\omega) = V_0 \int_{-\infty}^{\infty} e^{-at^2} \cos(\omega t) dt - iV_0 \int_{-\infty}^{\infty} e^{-at^2} \sin(\omega t) dt$$

Since e^{-at^2} is even and $\sin(\omega t)$ is odd, the right integral must be zero. The left integral evaluates to:

$$V_{in}(\omega) = V_0 \sqrt{\frac{\pi}{a}} e^{-\omega^2/(4a)}$$

We note that the Fourier transform of our Gaussian $V_{in}(t)$ yields a Gaussian $V_{in}(\omega)$. It is a special property of Gaussian functions that their Fourier transform is also a Gaussian function.

Next we multiply $V_{in}(\omega)$ by the magnitude of the transfer function to get $V_{out}(\omega)$. (We will ignore phase.) We found the magnitude of the transfer function for the RL low-pass filter in [section 3.2](#):

$$|H| = \frac{1}{\sqrt{1 + (\omega L/R)^2}}$$

The next step is to take the reverse Fourier transform to get $V_{out}(t)$. Finding an analytical expression for the reverse Fourier transform of $V_{out}(\omega) = V_{in}(\omega)|H|$ is difficult. (Feel free to give it a try.) But it is relatively straightforward to find $V_{out}(t)$ numerically using the FFT algorithm that is built into most data analysis software. To illustrate this, [Figure 3.6](#) shows both $V_{in}(t)$ as well as $V_{out}(t)$ found numerically, using the values $V_0 = 1.0$ V, $a = 1.0 \times 10^6$ s⁻², $R = 10$ Ω , and $L = 1.0$ mH. We see that the effect of the low-pass filter is to broaden our Gaussian peak.

3.4 TRANSFORMERS

A transformer is a device containing two sets of wire loops arranged such that the magnetic field lines produced by one set of loops will all (or nearly all) pass through the other set of loops. An AC current applied to one side of the transformer produces

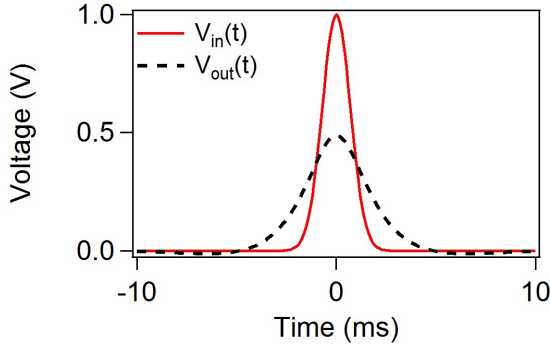


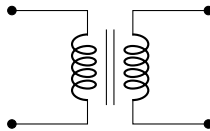
Figure 3.6 Gaussian voltage pulse before and after passing through an RL low-pass filter with a 3 dB point at $\omega = R/L = 10^4 \text{ s}^{-1}$.

a time-dependent magnetic field through the loops on that side. This results in a time-dependent magnetic flux in the other side which, from Faraday's law, produces an AC current at the same frequency as the signal applied to the first side. The transformer equation relates the currents and voltages on each side of the transformer. Its derivation can be found in most introductory physics textbooks, so here will simply state the result:

$$\frac{N_1}{N_2} = \frac{V_1}{V_2} = \frac{I_2}{I_1} \quad (3.11)$$

where N is the number of turns (loops) on each side of the transformer and V and I are the AC voltage and current, respectively, on each side. This equation assumes that the transformer is lossless and that all the magnetic field produced by one set of loops passes through the other set of loops. We see that the voltage ratio is proportional to the turns ratio and the current ratio is inversely proportional to the turns ratio. A transformer in which the output voltage is larger than the input voltage is called a step-up transformer, and one in which the output voltage is smaller than the input voltage is called a step-down transformer.

Assuming no loss, the power coming into the transformer input must be equal to the power coming out of the transformer output, i.e. $I_1 V_1 = I_2 V_2$. If side 1 is the input and side 2 is the output, then we can write the transfer function for our transformer as $H = V_2/V_1 = N_2/N_1$. We can represent a transformer in our circuit diagrams as:



The impedance on side 1 of the transformer can be written as $Z_1 = V_1/I_1$. Substituting from our transformer equation $V_1 = V_2(N_1/N_2)$ and $I_1 = I_2(N_2/N_1)$ and recognizing

that $V_2/I_2 = Z_2$ gives:

$$Z_1 = Z_2 \left(\frac{N_1}{N_2} \right)^2 \quad (3.12)$$

Thus we see that the impedance transforms as the turns ratio squared.

Transformers are often made by winding the wire loops around a ferrite core, which is a material with a high magnetic permeability and a low electrical conductivity. The high permeability ensures strong confinement of the magnetic field lines and boosts the coil inductance, while the low conductivity minimizes losses due to eddy currents.

Transformers are a critical part of our electric power grid. Specifically, AC power can be sent long distances from the power plant to our neighborhood at high voltage and small current, which minimizes the power dissipation due to resistive losses in the power lines. Then a transformer can be used to step down the voltage and step up the current for use in our house. An example illustrating the benefit of this approach is done in [Exercise 3.2](#).

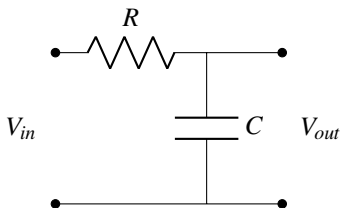
Exercise 3.2 Resistive Losses in Power Lines

Consider an electrical power line with a total length $l = 100$ km that is made of copper wire with a radius $r = 1.5$ mm and a resistivity $\rho = 1.68 \times 10^{-8}$ Ωm . Calculate the power dissipated due to the line resistance if the AC signal on the power line has (a) an rms voltage of 120 V and an rms current of 15 A, and (b) an rms voltage of 12,000 V and an rms current of 0.15 A. (Note that, in both cases, the product of the voltage and the current, i.e. the power, is the same. In the second case, a transformer could be used to step down the voltage and step up the current by a factor of 100 at the point of use.)

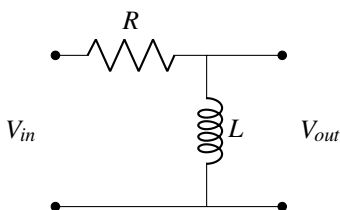
Solution: First we must find the total resistance of our power line. The cross-sectional area of our line is $A = \pi r^2 = 7.07 \times 10^{-6}$ m^2 . Then the resistance $R = \rho l / A = 237.7$ Ω . The power dissipated due to this resistance is $P = I_{\text{rms}}^2 R$. For case (a), $I_{\text{rms}} = 15$ A, and hence $P = 53,500$ W. For case (b), $I_{\text{rms}} = 0.15$ A, and hence $P = 5.35$ W. We see that case (a) results in $10,000 \times$ greater losses than case (b)!

3.5 PROBLEMS

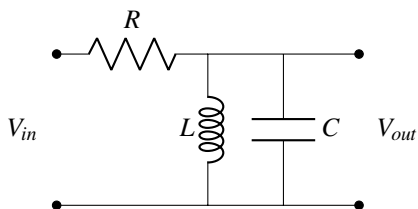
Problem 1. For the four-terminal circuit shown, find an expression for the transfer function $H(\omega)$ and its magnitude $|H(\omega)|$ in terms of the given circuit parameters. Be sure to simplify your answers. Then specify whether this is a low-pass, high-pass, band-pass, or band-stop filter.



Problem 2. For the four-terminal circuit shown, find an expression for the transfer function $H(\omega)$ and its magnitude $|H(\omega)|$ in terms of the given circuit parameters. Be sure to simplify your answers. Then specify whether this is a low-pass, high-pass, band-pass, or band-stop filter.



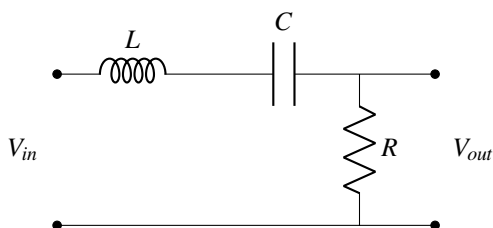
Problem 3. For the four-terminal circuit shown, find an expression for the transfer function $H(\omega)$ and its magnitude $|H(\omega)|$ in terms of the given circuit parameters. Be sure to simplify your answers. Then specify whether this is a low-pass, high-pass, band-pass, or band-stop filter.



Problem 4. The circuit below is a band-pass filter whose transfer function magnitude is given by

$$|H| = \frac{1}{\sqrt{1 + \frac{1}{R^2} \left(\omega L - \frac{1}{\omega C} \right)^2}}$$

If $R = 100 \, \Omega$, $L = 5.0 \, \mu\text{H}$, and $C = 8.0 \, \text{nF}$, what is the bandwidth Δf of the filter's pass-band? Note that the pass-band is defined as the range of frequencies for which $|H| \geq 1/\sqrt{2}$.



Problem 5. $V(t)$ is equal to 1.0 V for all time $|t| \leq 0.50$ s and is equal to zero for $|t| > 0.50$ s. Find an expression for $V(\omega)$, the Fourier transform of $V(t)$. Be sure to simplify your answer.

Problem 6. Consider a Gaussian voltage pulse given by $V(t) = V_0 e^{-\alpha t^2}$ where $\alpha = 1.0 \times 10^9 \text{ s}^{-2}$. Beyond what linear frequency f does the corresponding frequency-domain signal $V(f)$ fall below $1/\sqrt{2}$ of its maximum value?

Problem 7. A $75 \, \Omega$ resistor and a $1.5 \, \mu\text{F}$ capacitor are connected in series across side 1 of an ideal transformer. If the transformer has a turns ratio $N_1/N_2 = 10$, what is the magnitude of the impedance measured on side 2 of the transformer at a frequency $f = 2.0 \text{ kHz}$?



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

4 Transmission Lines

In DC circuits, the voltage at all points along an ideal wire is constant because an ideal wire has zero resistance. In AC circuits, we can treat the voltage along an ideal wire as constant at each moment in time as long as the length of the wire is much less than the signal wavelength. If the length of the wire is not much less than the signal wavelength, however, then we have to account for the spatial variation of the signal along the length of our wire. In this regime, we have to start thinking about our wire as something called a transmission line.

The signal wavelength on a wire is typically the same order of magnitude as the free space wavelength $\lambda_{fs} = c/f$, where c is the speed of light in free space and f is the signal frequency. At a frequency of 1 kHz, the free space wavelength is 300,000 m; this is a great deal longer than the wires in a typical circuit. At 1 MHz, the free space wavelength is 300 m; this is still much longer than the wires in most circuits. At 1 GHz, the free space wavelength is 0.3 m; this may not be much longer than the wires in a circuit, and hence we may need to think of our wires as transmission lines in order to understand the behavior of our circuit at this frequency.

All transmission lines have some capacitance per unit length and some inductance per unit length. Where the capacitance and inductance come from depends on the geometry, and there are many different transmission line geometries. One common geometry is the coaxial cable, which consists of a wire with a circular cross-section surrounded by a uniform thickness of an insulating material (typically a plastic such as polyethylene) which is in turn surrounded by a conducting layer (often a copper braid). The inductance is dominated by the self-inductance of the center wire, and the capacitance is between the center wire and the outer conductor.

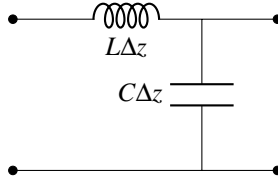
Another transmission line geometry is the twin wire, which consists of a pair of wires with a uniform spacing. The microstrip is a transmission line geometry that is commonly used on circuit boards. It consists of a thin strip of metal of some width on one side of the insulating circuit board with a continuous metal coating on the opposite side of the circuit board. A single-wire can also be a transmission line. In this case, the capacitance comes not from two separate conductors but rather from one part of the wire to another part of the same wire. A single-wire transmission line is sometimes called a Sommerfeld-Goubau line.

4.1 LUMPED-ELEMENT MODEL

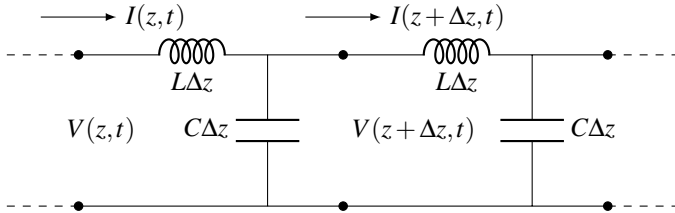
In this section, we will develop a model to understand how the inductance per unit length and capacitance per unit length of a transmission line determine its behavior. We will assume that the transmission line has a center conductor with inductance per unit length L (with units of H/m) along with a capacitance per unit length between the

center conductor and the outer conductor C (with units of F/m). We will assume there is no loss. Of course, all transmission lines have some loss, but this is a reasonable approximation to the real case as long as the losses are sufficiently small.¹

We begin by constructing a unit cell consisting of a series inductance and a capacitance to ground, where ground is represented by an ideal wire. We call the length of this unit cell Δz . Since L and C are defined per unit length, the inductance of our unit cell is $L\Delta z$ and the capacitance of our unit cell is $C\Delta z$:



A lumped-element component is one whose dimensions are all much smaller than the signal wavelength. Here we assume that Δz is much smaller than the signal wavelength so that our unit cell can be treated as a lumped element. We then construct our transmission line, whose length is not much smaller than the signal wavelength, by repeating our unit cell many times:



We call the voltage between the center conductor and the outer conductor at the input of the left-most unit cell shown $V(z, t)$. The voltage at the output of that unit cell, and hence at the input of the next unit cell, is then $V(z + \Delta z, t)$. We call the current flowing into the inductor of the left-most unit cell $I(z, t)$. Then the current flowing into the inductor of the next unit cell is $I(z + \Delta z, t)$.

We then use Kirchhoff's rules to write two equations. The first comes from applying the loop rule around our left-most unit cell:

$$V(z, t) - L\Delta z \frac{\partial I(z, t)}{\partial t} - V(z + \Delta z, t) = 0 \quad (4.1)$$

The second comes from applying the junction rule at the point where the top conductors of the two unit cells connect:

$$I(z, t) - C\Delta z \frac{\partial V(z + \Delta z, t)}{\partial t} - I(z + \Delta z, t) = 0 \quad (4.2)$$

¹For a derivation including losses, see for example [chapter 2](#) of *Microwave Engineering* by D.M. Pozar.

Dividing both equations by Δz and rearranging, we get:

$$\frac{V(z + \Delta z, t) - V(z, t)}{\Delta z} = -L \frac{\partial I(z, t)}{\partial t}$$

$$\frac{I(z + \Delta z, t) - I(z, t)}{\Delta z} = -C \frac{\partial V(z + \Delta z)}{\partial t}$$

Taking the limit $\Delta z \rightarrow 0$, we go from finite differences in z to partial derivatives:

$$\frac{\partial V}{\partial z} = -L \frac{\partial I}{\partial t} \quad (4.3)$$

$$\frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t} \quad (4.4)$$

Taking the partial derivative with respect to z of both sides gives:

$$\frac{\partial^2 V}{\partial z^2} = -L \frac{\partial}{\partial t} \frac{\partial I}{\partial z}$$

$$\frac{\partial^2 I}{\partial z^2} = -C \frac{\partial}{\partial t} \frac{\partial V}{\partial z}$$

Here we have made use of the fact that our partial derivatives with respect to position and time commute.

We can then plug in our previous pair of equations for $\partial V/\partial z$ and $\partial I/\partial z$ (Equations 4.3 and 4.4) to get:

$$\frac{\partial^2 V}{\partial z^2} = LC \frac{\partial^2 V}{\partial t^2}$$

$$\frac{\partial^2 I}{\partial z^2} = LC \frac{\partial^2 I}{\partial t^2}$$

If we assume that the time-dependence of the voltage and current is a simple oscillatory function, we can plug the expressions $V(t) = V_0 e^{i\omega t}$ and $I(t) = I_0 e^{i\omega t}$ into the right-hand side of our expressions to get:

$$\frac{\partial^2 V}{\partial z^2} = -\omega^2 LC V(z) \quad (4.5)$$

$$\frac{\partial^2 I}{\partial z^2} = -\omega^2 LC I(z) \quad (4.6)$$

These equations have the form of a one-dimensional spatial wave equation. A general form of such a wave equation can be written as

$$\frac{\partial^2 y}{\partial x^2} = \gamma^2 y(x) \quad (4.7)$$

where $y(x)$ is the wave amplitude at position x and γ is known as the propagation constant, which is in general a complex number. The solutions to our wave equation are of the form

$$y(x) = y_0^+ e^{-\gamma x} + y_0^- e^{\gamma x} \quad (4.8)$$

where y_0^+ denotes the amplitude of the wave moving in the $+x$ direction and y_0^- denotes the amplitude of the wave moving in the $-x$ direction. γ is often written in the form $\gamma = \alpha + i\beta$, where α is called the attenuation constant and β is called the phase constant. The phase velocity of our wave is $v = \omega/\beta$ and the wavelength is $\lambda = 2\pi/\beta$.

By comparing Equations 4.5 and 4.6 to our general form of the wave equation (Equation 4.7), we can write a solution for the position-dependent voltage

$$V(z) = V_0^+ e^{-\gamma z} + V_0^- e^{\gamma z} \quad (4.9)$$

and a solution for the position-dependent current

$$I(z) = I_0^+ e^{-\gamma z} + I_0^- e^{\gamma z} \quad (4.10)$$

with a propagation constant $\gamma = i\omega\sqrt{LC}$. We see that here the propagation constant is purely imaginary, i.e. the phase constant $\beta = \omega\sqrt{LC}$ and the attenuation constant $\alpha = 0$. This means that the current and voltage have a wavelength

$$\lambda = 2\pi/(\omega\sqrt{LC}) \quad (4.11)$$

and a phase velocity

$$v = 1/\sqrt{LC} \quad (4.12)$$

and their amplitude does not decay as the wave travels down the transmission line. If our model included loss, we would get a nonzero α and the wave amplitude would decay by a factor of $1/e$ each time it travels a distance of $1/\alpha$ along the transmission line.

Taking the spatial derivative of our Equations 4.9 and 4.10 gives:

$$\frac{\partial V}{\partial z} = -\gamma V(z) \quad (4.13)$$

$$\frac{\partial I}{\partial z} = -\gamma I(z) \quad (4.14)$$

If we plug our oscillatory time-dependent solutions for $V(t)$ and $I(t)$ into Equations 4.3 and 4.4, we get:

$$\frac{\partial V}{\partial z} = -i\omega LI(z, t) \quad (4.15)$$

$$\frac{\partial I}{\partial z} = -i\omega CV(z, t) \quad (4.16)$$

We can then combine Equations 4.13 and 4.15 and Equations 4.14 and 4.16:

$$\gamma V(z, t) = i\omega LI(z, t) \quad (4.17)$$

$$\gamma I(z, t) = i\omega CV(z, t) \quad (4.18)$$

The impedance of our transmission line, $Z_0 = V(z, t)/I(z, t)$ can then be found from either Equation 4.17 or 4.18, which, using $\gamma = i\omega\sqrt{LC}$, gives the same result in either case:

$$Z_0 = \sqrt{\frac{L}{C}} \quad (4.19)$$

We call this the *characteristic impedance* of the transmission line. Note that, while our model was constructed entirely from elements with purely imaginary impedances, the characteristic impedance is a purely real number. This does not mean that it dissipates power; our model has no loss, so it dissipates zero power. The characteristic impedance is simply the ratio of the voltage $V(z, t)$ to the current $I(z, t)$, and the fact that it is purely real means that the voltage and current are in-phase with each other as the signal travels down the transmission line.

We see that, for our lossless transmission line, if we know the inductance per unit length L and the capacitance per unit length C , we can determine the signal wavelength (Equation 4.11), the phase velocity (Equation 4.12), and the characteristic impedance (Equation 4.19).

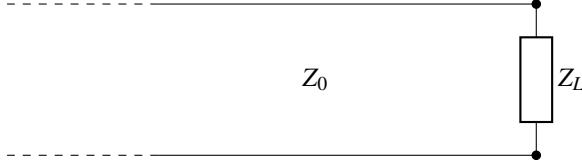
Exercise 4.1 Properties of RG-58/U Coaxial Cable

RG-58/U is a type of coaxial cable with a center conductor diameter of 0.8 mm surrounded by a polyethylene dielectric of thickness 1.1 mm which itself is surrounded by a braided copper sleeve and a protective plastic outer jacket. The relative dielectric constant of polyethylene is $\epsilon_r = 2.3$. Assuming that losses are negligible, what are the phase velocity and characteristic impedance of the RG-58/U coaxial cable?

Solution: To find the inductance per unit length L and capacitance per unit length C , we need to revisit some results from electromagnetism. The inductance per unit length of a coaxial cable is given by $L = \mu_0/(2\pi) \ln(b/a)$ H/m, where $\mu_0 = 4\pi \times 10^{-7}$ H/m is the permeability of free space, b is the inside diameter of our copper sleeve, and a is the diameter of our center conductor. Using $b = 3.0$ mm and $a = 0.8$ mm, we get $L = 264$ nH/m. The capacitance per unit length of a coaxial cable is given by $C = 2\pi\epsilon_0\epsilon_r/\ln(b/a)$ F/m, where $\epsilon_0 = 8.85 \times 10^{-12}$ F/m is the permittivity of free space. For our cable, this corresponds to $C = 97$ pF/m. Using these values of L and C in Equation 4.12, we get a phase velocity $v = 2.0 \times 10^8$ m/s, or two-thirds the speed of light in free space. And from Equation 4.19, we get a characteristic impedance $Z_0 = 52 \Omega$.

4.2 TERMINATED TRANSMISSION LINES

We will now consider a lossless transmission line with characteristic impedance Z_0 that is terminated by a load impedance Z_L . We will assume Z_L is a lumped-element component, i.e. it is much smaller than the signal wavelength and can be treated as localized at one point in space.



We will define the position of Z_L as $z = 0$. As in the previous section, we can write traveling-wave expressions for the voltage on our transmission line $V(z)$ and the current on our transmission line $I(z)$:

$$V(z) = V_0^+ e^{-\gamma z} + V_0^- e^{\gamma z} \quad (4.20)$$

$$I(z) = I_0^+ e^{-\gamma z} + I_0^- e^{\gamma z} \quad (4.21)$$

where the $+$ superscript denotes the right-moving amplitude, the $-$ superscript denotes the left-moving amplitude, and $\gamma = i\omega\sqrt{LC}$. Using $Z_0 = V(z)/I(z)$, we can express the current on our transmission line as

$$I(z) = \frac{V_0^+}{Z_0} e^{-i\gamma z} - \frac{V_0^-}{Z_0} e^{\gamma z} \quad (4.22)$$

where the negative sign on the second term reflects the fact that it is a left-moving current, i.e. the current is flowing in the negative- z direction.

At $z = 0$, the impedance Z of our transmission line must be equal to Z_L . Thus we can write:

$$Z_L = Z(0) = \frac{V(0)}{I(0)} = \frac{V_0^+ + V_0^-}{\frac{V_0^+}{Z_0} - \frac{V_0^-}{Z_0}} = Z_0 \frac{V_0^+ + V_0^-}{V_0^+ - V_0^-}$$

where the voltage amplitudes V_0^+ and V_0^- in this expression are defined at the position $z = 0$. This can be algebraically rearranged as:

$$V_0^- = V_0^+ \frac{Z_L - Z_0}{Z_L + Z_0}$$

We define the *reflection coefficient* Γ as the ratio of the voltage reflected off the load – i.e. V_0^- , the left-moving amplitude at $z = 0$ – to the voltage incident on the load – i.e. V_0^+ , the right-moving amplitude at $z = 0$ – and hence we have:

$$\Gamma = \frac{V_0^-}{V_0^+} = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (4.23)$$

Γ is in general a complex number, reflecting the fact that the incident and reflected voltages are not necessarily in-phase. The ratio of power reflected off Z_L to the power incident on Z_L is equal to Γ multiplied by its complex conjugate, $|\Gamma|^2$. Because of conservation of energy, whatever incident power isn't reflected must be absorbed, and hence the ratio of the power absorbed by Z_L to the power incident on Z_L is $1 - |\Gamma|^2$.

Exercise 4.2 Finding the Reflection Coefficient

A transmission line with characteristic impedance $Z_0 = 75 \, \Omega$ is terminated by a lumped-element load consisting of a 25 pF capacitor in parallel with a 100 Ω resistor. At a frequency $f = 50$ MHz, what are the reflection coefficient and the fraction of reflected power?

Solution: First we need to find Z_L , which is the parallel combination of the resistor impedance and the capacitor impedance, i.e.

$$Z_L = \frac{1}{1/R + i2\pi fC} = \frac{1/R - i2\pi fC}{(1/R)^2 + (2\pi fC)^2} = 61.8 - i48.6 \, \Omega$$

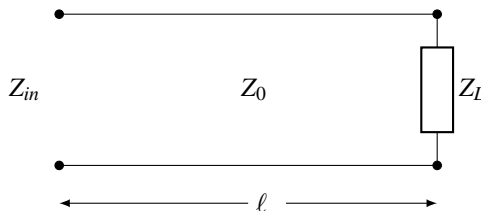
The reflection coefficient can then be found from Equation 4.23:

$$\Gamma = 0.0266 - i0.346$$

And the fraction of reflected power is:

$$|\Gamma|^2 = 0.120$$

Next we will consider the impedance Z_{in} that is measured at the input of a transmission line with characteristic impedance Z_0 and length ℓ that is terminated at the opposite end by an impedance Z_L :



Using our definition of Γ , we can rewrite Equations 4.20 and 4.22 as:

$$V(z) = V_0^+ (e^{-\gamma z} + \Gamma e^{\gamma z}) \quad (4.24)$$

$$I(z) = \frac{V_0^+}{Z_0} (e^{-\gamma z} - \Gamma e^{\gamma z}) \quad (4.25)$$

Defining the location of Z_L as $z = 0$, the input impedance Z_{in} is then the impedance of our transmission line at the position $z = -\ell$:

$$Z_{in} = \frac{V(-\ell)}{I(-\ell)} = Z_0 \frac{e^{-\gamma\ell} + \Gamma e^{\gamma\ell}}{e^{-\gamma\ell} - \Gamma e^{\gamma\ell}} = Z_0 \frac{1 + \Gamma e^{-2\gamma\ell}}{1 - \Gamma e^{-2\gamma\ell}}$$

Using Euler's formula and some manipulation, this can be expressed as

$$Z_{in} = Z_0 \frac{Z_L + iZ_0 \tan(\beta\ell)}{Z_0 + iZ_L \tan(\beta\ell)} \quad (4.26)$$

where we have also used the fact that $\gamma = i\beta$ for the lossless case considered here.

Recall that $\beta = 2\pi/\lambda$, and hence $\tan(\beta\ell) = \tan(2\pi\ell/\lambda)$. When the transmission line length $\ell = n\lambda/4$ with $n = 2, 4, 6, \dots$, then $\tan(2\pi\ell/\lambda) = 0$ and hence $Z_{in} = Z_L$. In this case, the system behaves as if the transmission line isn't there. And when $\ell = n\lambda/4$ with $n = 1, 3, 5, \dots$, then $\tan(2\pi\ell/\lambda) = \pm\infty$ and hence $Z_{in} = Z_0^2/Z_L$. This case is called a quarter-wave transformer, since the transmission line characteristic impedance Z_0 can be chosen in order to transform Z_L to a particular valued of Z_{in} . For purely real Z_L and Z_{in} , it is always possible to find a value of Z_0 that achieves the desired transformation, but for a complex Z_L and/or Z_{in} , it is not always possible.

Exercise 4.3 The Quarter-Wave Transformer

We want to couple a $75 \, \Omega$ load impedance to a transmission line with a $50 \, \Omega$ characteristic impedance with zero reflections at a frequency of 2.5 GHz using a quarter-wave transformer. What are the characteristic impedance and shortest possible length of the section of transmission line that comprises the quarter-wave transformer? Assume that the phase velocity of the quarter-wave transformer is two thirds the speed of light in free space.

Solution: In free space, a frequency of 2.5 GHz corresponds to a wavelength $\lambda_{fs} = c/f = 0.12 \, \text{m}$, where $c = 3.00 \times 10^8 \, \text{m/s}$ is the speed of light in free space. If the phase velocity is two thirds the speed of light in free space, then the wavelength is also two thirds of the free space value, $\lambda = 2\lambda_{fs}/3 = 0.080 \, \text{m}$. Our shortest quarter-wave condition ($n = 1$) is $\ell = \lambda/4 = 0.020 \, \text{m} = 2.0 \, \text{cm}$. At this length, we have $Z_{in} = Z_0^2/Z_L$. Solving for Z_0 gives $Z_0 = \sqrt{Z_{in}Z_L}$. Here we have $Z_L = 75 \, \Omega$ and we want $Z_{in} = 50 \, \Omega$ to match our $50 \, \Omega$ transmission line, and hence the characteristic impedance of our 2.0 cm long quarter-wave transformer should be $Z_0 = 61.2 \, \Omega$.

4.3 SPECIAL TOPIC: SCATTERING PARAMETERS

Scattering parameters, often called S-parameters, are conceptually related to the transfer function that was introduced in [chapter 3](#). The transfer function applies to

a four-terminal circuit with an input provided by a quasi-ideal voltage source (approximately zero internal resistance) and whose output is measured by a quasi-ideal voltmeter (approximately infinite internal resistance). For microwave frequency circuits, as we have seen in this chapter, it is more common to have the same input and output impedances, which ensures that there are no reflections at the points of connection and which results in maximal power transfer.

In this impedance-matched case, we can consider a multi-terminal circuit where each pair of terminals is connected to the same impedance, most commonly $50\ \Omega$. For microwave circuits, it is common to refer to each pair of terminals as a *port*. The scattering parameter S_{ba} is then defined as the ratio of the signal voltage going into port a to the signal voltage coming out of port b . Note that it is possible for a and b to be the same port. This means that, for a two-port circuit, there are four S-parameters: S_{11} , S_{21} , S_{12} , and S_{22} . They can be written in the form of a 2×2 matrix, known as a scattering matrix, which satisfies the matrix equation

$$\begin{bmatrix} V_1^{out} \\ V_2^{out} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} V_1^{in} \\ V_2^{in} \end{bmatrix}$$

where V_a^{in} is the voltage sent into port a and V_b^{out} is the voltage coupled out of port b . Since these voltages are not necessarily in phase, each S-parameter is in general a complex number, where the magnitude corresponds to the amplitude ratio of the two voltages and the phase angle corresponds to their phase difference. This approach can be generalized to describe an N -port circuit:

$$\begin{bmatrix} V_1^{out} \\ V_2^{out} \\ \dots \\ V_N^{out} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1N} \\ S_{21} & & & \dots \\ \dots & & & \\ S_{N1} & \dots & & S_{NN} \end{bmatrix} \begin{bmatrix} V_1^{in} \\ V_2^{in} \\ \dots \\ V_N^{in} \end{bmatrix}$$

It is important to remember that this assumes that all ports are terminated by the same (usually $50\ \Omega$) impedance. The scattering parameter S_{ab} for $a \neq b$ is sometimes called the transmission coefficient. If all ports are terminated by a matched impedance, then the reflection coefficient Γ measured at port a is equivalent to the scattering parameter S_{aa} .

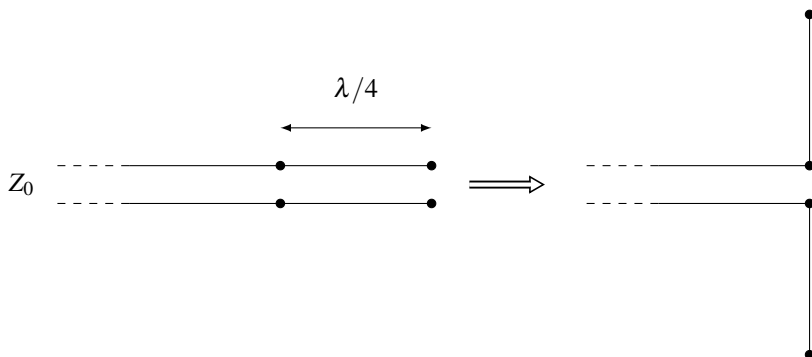
4.4 SPECIAL TOPIC: THE HALF-WAVE DIPOLE ANTENNA

Free space has a characteristic impedance $Z_{fs} = \sqrt{\mu_0/\epsilon_0} \approx 377\ \Omega$, where $\mu_0 = 4\pi \times 10^{-7}\ \text{H/m}$ is the permeability of free space and $\epsilon_0 = 8.85 \times 10^{-12}\ \text{F/m}$ is the permittivity of free space. Electromagnetic waves travel in free space at a speed $c = 1/\sqrt{\mu_0\epsilon_0} \approx 3.00 \times 10^8\ \text{m/s}$. Note the similarity between these expressions and those for the characteristic impedance and phase velocity of our lossless transmission line. Of course, the signal in a transmission line consists of a coupled oscillating voltage and current, while an electromagnetic wave consists of a coupled oscillating electric and magnetic field.

An oscillating current and voltage in a circuit can produce an electromagnetic wave, and an electromagnetic wave can produce an oscillating current and voltage in a circuit. A device that is designed to do these things efficiently is called an antenna. There are many different types of antennas. Here we will consider one type, the half-wave dipole antenna.

A dipole antenna consists of two lengths of wire extending in different – usually opposite – directions. In the middle of the antenna, known as the feed, each wire connects to a transmission line or some other electrical circuit. The electrical circuit can drive oscillating currents in the dipole wires, which couple to electromagnetic waves in free space. Similarly, if an electromagnetic wave is incident on the antenna, it will drive an oscillating current in the wires.

A dipole antenna has a maximum in its efficiency when the total length of the two wires is approximately equal to half the signal wavelength. Such a half-wave dipole can be thought of as similar to a quarter wave transformer, except instead of matching a transmission line impedance to some load impedance, it is matching a transmission line impedance to the impedance of free space:



When the total dipole length is $\lambda/2$, there will be a half-wave resonance along the length of the dipole. The far ends of the dipole are not electrically connected to anything and so they impose a zero of the current, and hence the current is a maximum in the center, where the feed is located.

For efficient operation, one wants the input impedance of the antenna to be close to the impedance of the transmission line that is connected to the antenna feed. Assuming the wire diameter is negligible compared to the wavelength, a dipole has an input impedance whose real part is given by $2P_{rad}/|I_0|^2$, where P_{rad} is the total radiated power and I_0 is the amplitude of the current at the feed. For a half-wave dipole in free space, this is equal to approximately 73Ω . When the antenna is lossless, the real part of the input impedance is equivalent to a quantity called the radiation resistance. 73Ω is conveniently close to the characteristic impedance of typical transmission lines, which are commonly 50Ω or 75Ω .

For maximum efficiency, one wants the imaginary part of the input impedance to be zero. The imaginary part of the input impedance goes to zero when the dipole length is very close, but not exactly equal, to $\lambda/2$; exactly how far off it is from

$\lambda/2$ depends on the wire diameter. In practice, half-wave dipoles are often made just slightly shorter than $\lambda/2$ to get the imaginary part of the input impedance as close to zero as possible.²

Antennas are characterized by, among other things, their radiation pattern. The radiation pattern describes the directional efficiency by which the antenna both emits and receives electromagnetic radiation. This equivalency of emission efficiency and reception efficiency is known as reciprocity. The radiation pattern is often visualized with a three-dimensional surface whose distance from the origin is proportional to the power emitted in that direction. For a dipole antenna, such a surface is the shape of a torus, as seen in [Figure 4.1](#), with the dipole oriented through the hole in the center of the torus with the feed at the origin. This reflects the fact that the emission efficiency is uniform along the radial direction of the dipole and goes to zero along the axial direction.

When we discuss radiation patterns, we are most commonly concerned with what is known as the far-field radiation pattern. This is the radiation pattern at a distance $d \geq 2L^2/\lambda$ from the antenna, where L is the largest dimension of the antenna and λ is the wavelength of the electromagnetic wave. $2L^2/\lambda$ is known as the Fraunhofer distance d_F . Distances less than d_F are in the near-field region. Distances between d_F and $0.62\sqrt{L^3/\lambda}$ are known as the radiating near-field or Fresnel region and distances less than $0.62\sqrt{L^3/\lambda}$ are known as the reactive near-field region. The electromagnetic interaction between an antenna and another object will differ depending on which of these regions the object is in.

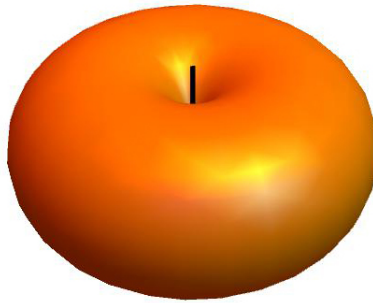


Figure 4.1 Radiation pattern of a dipole antenna.

The two conductors in a dipole antenna are generally symmetric, while most transmission lines are not; transmission lines are comprised of conductors with different geometries, such as the center conductor and outer conductor of a coaxial cable. This difference can generate reflections at the point of connection between the transmission line and the antenna. These reflections can be minimized with a device called a *balun*. The name balun is a portmanteau of the words *balanced* and

²For details, see for example the textbook *Antenna Theory* by C.A. Balanis.

unbalanced. A balun converts between an asymmetric or unbalanced signal and a symmetric or balanced signal.

We will be leaving our discussion of antennas here and moving on to other topics. There are a number of textbooks dedicated to antennas where the interested reader can find much more information beyond what we have covered in this very brief introduction.

4.5 PROBLEMS

Problem 1. RG6 is a type of coaxial cable with an inductance per unit length of 380 nH/m and a capacitance per unit length of 66 pF/m. (a) Assuming losses are small, what are the characteristic impedance and phase velocity of the cable? (b) A voltage signal in this cable has an initial rms amplitude of 2.0 V. After the signal travels 10 m down the length of the cable, its rms amplitude is now 1.7 V. What is the value of α , the real part of the cable's propagation constant, at the signal frequency?

Problem 2. A lossless transmission line with characteristic impedance $Z_0 = 50 \Omega$ is terminated by a lumped-element load consisting of a 35 pF capacitor in parallel with a 100 Ω resistor. At a frequency $f = 45$ MHz, what are the reflection coefficient and the fraction of reflected power?

Problem 3. A lossless transmission line of length $l = 0.75\lambda$ has a characteristic impedance $Z_0 = 50 \Omega$ and is terminated with a load impedance $Z_L = 45 - i30 \Omega$. Find the impedance seen at the input of the transmission line Z_{in} . Express your answer in the form $Z_{in} = X + iY$, i.e. separate the real and imaginary parts.

Problem 4. A lossless transmission line is terminated with a 110 Ω load. The voltage standing wave ratio (VSWR) is defined as the ratio of the maximum voltage amplitude on the line to the minimum voltage amplitude on the line and can be expressed as

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|}$$

where Γ is the reflection coefficient. If the VSWR on the line is 1.55, find the two possible values of the characteristic impedance of the transmission line.

5 Semiconductor Devices

5.1 PHYSICS OF SEMICONDUCTORS

A crystal is a solid whose atoms are arranged in a periodic structure. In a crystal, the positive ions consisting of the atomic nuclei and inner-shell electrons create a periodic potential energy. To understand how a mobile electron will move through the crystal, we can solve the Schrodinger equation for an electron in this periodic potential. The solution will yield the allowed electron energies. A range of energies for which there are allowed electron states is known as a *band*, and a range of energies for which there are no allowed electron states is known as a *bandgap*. Performing such a calculation is not trivial, and the details are best left for a course in solid state physics or solid state chemistry. In many cases, exact analytical solutions are impossible, so a numerical procedure is used.

Electrons obey the Pauli exclusion principle, which says that two electrons in a system cannot have identical sets of quantum numbers. Each allowed energy state is characterized by some set (or sets) of quantum numbers. At zero temperature, all the electrons in a crystal will sit in the lowest available energy state that does not violate the Pauli exclusion principle. The greater the density of these electrons, the higher the highest filled energy state. We call this highest filled energy state at zero temperature the *Fermi energy* (E_F). At finite temperature, the Fermi energy is replaced by the *chemical potential* $\mu(T)$, with $\mu(0) = E_F$.¹

If no extra energy has been added to the system, then all allowed electron states below E_F will be occupied and all allowed electron states above E_F will be unoccupied. If the Fermi energy lies in the middle of a band, then the addition of a small amount of energy to the system will excite an electron from a filled state into an available empty state. We call such a material a *metal*. In a metal, if we apply a small voltage across the material, a current will flow. This is because a voltage V adds an energy q_0V , where $q_0 = 1.602 \times 10^{-19}$ C is the fundamental charge. Regardless of the value of V , in a metal this voltage will be able to excite electrons from below E_F to above E_F , and these excited electrons will move in response to the electric field associated with the voltage, resulting in a current. The number of electrons that are excited depends on both the electron density and the voltage; a higher electron density and a larger voltage will both correspond to a larger current. The value of the current also depends on the scattering rate of electrons, as we saw in [section 1.1](#).

¹In the semiconductor literature, the term *Fermi energy* is sometimes treated as synonymous with the term *chemical potential*, even though this is only technically true at zero temperature. Because room temperature is very small compared to typical values of the Fermi energy (expressed in equivalent units), treating the Fermi energy and the chemical potential as the same at room temperature is often a reasonable approximation.

If the Fermi energy falls in a bandgap, then the material is an *insulator*. In this case, at zero temperature, all the available electron states below the bandgap are filled, and all the available electron states above the bandgap are empty. The band below E_F is known as the *valence band*, and the band above E_F is known as the *conduction band*. In order to excite an electron from a filled state in the valence band to an empty state in the conduction band, there is a minimum required energy that is equal to the bandgap energy E_g . If we apply a voltage V with $q_0V < E_g$, no electrons will be excited and no current will flow. A *semiconductor* is an insulator for which the bandgap energy is not too large, typically between a few electron volts (eV) and a few tenths of an eV. When discussing semiconductors, it is common to specify energies in units of eV, even though it is not an SI unit. 1 eV is defined as the kinetic energy gained by an electron that has been accelerated across an electric potential difference of 1 V, and hence $1 \text{ eV} = 1.602 \times 10^{-19} \text{ joules (J)}$. Since joules are the SI unit of energy, it is usually a good idea to convert from electron volts to joules before performing calculations.

There are different ways to add energy to the electrons in a crystal. One is by adding thermal energy; a temperature T corresponds to an average thermal energy of approximately $k_B T$, where $k_B = 1.381 \times 10^{-23} \text{ J/K}$ is the Boltzmann constant. Note that at room temperature ($T = 293 \text{ K}$), $k_B T = 0.025 \text{ eV}$, which is very small compared to most semiconductor bandgap energies. However, it is important to remember that temperature is a statistical quantity; if the average energy of the electron system is $k_B T$, some electrons will have a greater energy and some will have a smaller energy. The thermal distribution of electron energies at temperature T is given by a Fermi-Dirac distribution, which says that the probability that an electron state at energy E will be occupied is given by:

$$N(E) = \frac{1}{e^{\frac{E-\mu}{k_B T}} + 1} \quad (5.1)$$

For a pure semiconductor, the energy level of the chemical potential μ is at or very close to the middle of the bandgap.² In the high-energy limit ($E \gg k_B T$ and $E \gtrsim 1.1E_F$), Equation 5.1 can be approximated as $N(E) \propto e^{-E/(k_B T)}$. Thus at room temperature, even though $k_B T \ll E_g$, we can still have a small amount of thermal promotion of electrons above the bandgap due to the exponentially decaying high-energy tail of the Fermi-Dirac distribution.

Another way to add energy is by applying a voltage V ; we already said that applying a voltage V corresponds to adding an energy q_0V , which means that 1 V corresponds to an energy of 1 eV. A third way to add energy is by absorbing light; each photon of light that is absorbed provides an energy of hf , where $h = 6.626 \times 10^{-34} \text{ J}\cdot\text{s}$ is Planck's constant and f is the frequency of the light. For visible light, $f \approx 5 \times 10^{14} \text{ Hz}$, and hence $hf \approx 2 \text{ eV}$; this value is conveniently close

²The energy level of the chemical potential in an intrinsic semiconductor is shifted relative to the middle of the bandgap by an energy of approximately $(3/4)k_B T \ln(m_h/m_e)$, where m_h is the hole effective mass and m_e is the electron effective mass.

to typical semiconductor bandgap energies. When a photon is absorbed, it gives all its energy to a single electron, but that electron can then share that energy with other electrons via inelastic scattering.

An energy diagram for a semiconductor at zero temperature is shown in Figure 5.1. The y-axis is energy, and the x-axis is used to indicate whether or not there are available electron states. At zero temperature, all states in the valence band are filled and all states in conduction band are empty. The chemical potential μ is located (approximately) in the middle of the bandgap.

Figure 5.2 illustrates the effect of absorbing a photon with frequency $f \geq E_g/h$. The photon promotes an electron from the valence band to the conduction band, leaving behind a vacancy in the valence band. The electron in the conduction band is now mobile. The vacancy in the valence band is known as a *hole* and behaves as a mobile particle with a charge $+q_0$. Thus the absorption of one photon creates two mobile charged particles, both of which will move in response to an applied electric field, resulting in a current. The resistance of the semiconductor is inversely proportional to the density of mobile charges, which includes both the density of electrons in the conduction band and the density of holes in the valence band. Specifically, we can write the electrical conductivity of the semiconductor as

$$\sigma = n_e q_0 \mu_e + n_h q_0 \mu_h \quad (5.2)$$

where n_e is the volume density of conduction-band electrons, n_h is the volume density of valence-band holes, $\mu_e = q_0 \tau_e / m_e$ is the electron mobility, and $\mu_h = q_0 \tau_h / m_h$ is the hole mobility. Here m_e is the electron effective mass, m_h is the hole effective mass, τ_e is the average electron scattering time, and τ_h is the average hole scattering time. (A discussion of effective masses is probably best left for a course in solid state physics.)

At room temperature, the thermal promotion of electrons from the valence band to the conduction band is very small, so most pure semiconductors have a very high room-temperature resistivity. For making electrical components, it is useful to be able to control the room-temperature resistivity of a semiconductor. This is accomplished through a process called *doping*. There are two types of doping: n-type and p-type.

In n-type doping, a small percentage of the original atoms in the semiconductor crystal are replaced with a type of dopant atom that has one or more extra outer-shell electrons. (It is most common for the dopant atoms to have one extra outer shell electron.) For doping to work, the dopant atoms should bond in the crystal lattice as if they were the original atom. This is more likely to work when the dopant atom is similar in size to the original atom and when the percentage of dopant atoms is small. When this happens, the extra outer-shell electrons end up in the conduction band. Thus one can adjust the density of conduction-band electrons by adjusting the density of dopant atoms, with each dopant atom donating its extra outer shell electron(s) to the conduction band. In our energy diagram, we draw a donor energy level inside the bandgap but very close to the edge of the conduction band. This donor level is where the positively-charged donor ions live. (They are positively charged because they have lost an electron.) Each immobile donor ion is associated with a mobile electron in the conduction band. This is shown schematically in Figure 5.3.

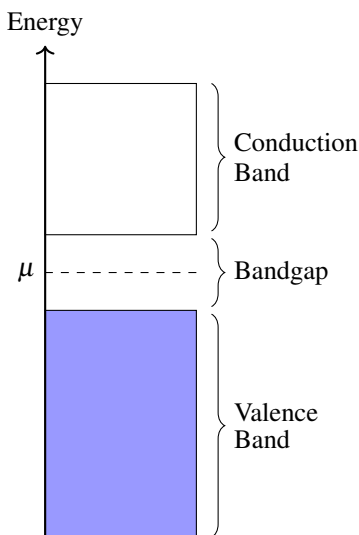


Figure 5.1 Energy diagram of a semiconductor at zero temperature. The shaded rectangle represents filled states and the empty rectangle represents empty states. These are separated by the bandgap, at the center of which is the chemical potential μ .

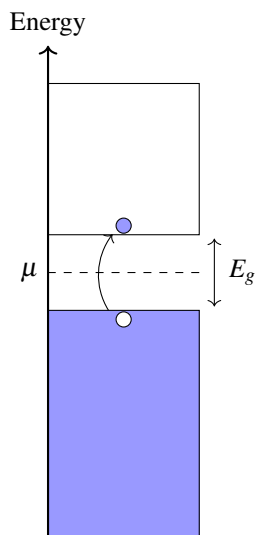


Figure 5.2 Energy diagram of a semiconductor showing the result of absorbing a photon with $hf \geq E_g$, which promotes one electron from the valence band to the conduction band and leaves behind a hole in the valence band.

P-type doping is similar to n-type doping except the dopant atoms contain at least one fewer outer-shell electron than the original atoms. (It is most common for the dopant atoms to have one fewer outer-shell electron.) In order to bond in the crystal lattice like the original atoms, the dopant atoms must take an electron from the valence band, which makes them negatively charged and leaves behind a mobile hole. In our energy diagram, we draw the negative dopant ions at the acceptor energy level, which is inside the bandgap but very close to the edge of the valence band. Each immobile acceptor ion is associated with a mobile hole in the valence band. This is shown schematically in Figure 5.4. It is important to remember that the acceptor and donor ions are not mobile, while the electrons in the conduction band and holes in the valence band are mobile. In a doped semiconductor, the chemical potential becomes effectively pinned at the donor or acceptor energy level.

Let's consider silicon (Si) as an example semiconductor. It is common to n-type dope silicon using phosphorous (P). Phosphorus is in the same row and one column to the right of silicon in the periodic table, so it has one extra outer-shell electron as well as a very similar size. It is common to p-type dope silicon using boron (B). Boron is one column to the left and one row above silicon in the periodic table, so it has one fewer outer-shell electron and a roughly similar size. One column to the

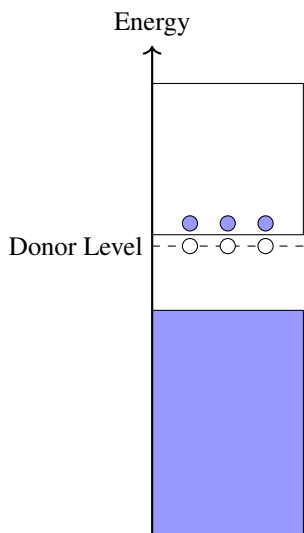


Figure 5.3 Energy diagram of an n-type doped semiconductor

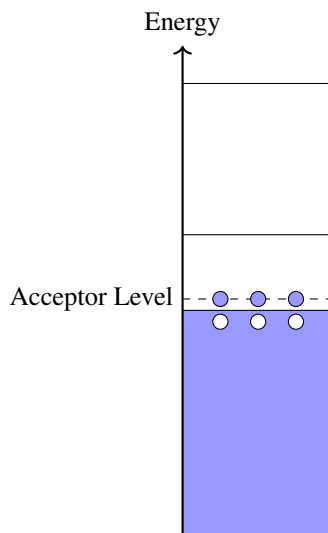


Figure 5.4 Energy diagram of a p-type doped semiconductor.

left of silicon in the same row is aluminum (Al), which is closer in size to silicon than boron. Aluminum can be used to dope silicon, but it is not as commonly used as boron; this is because boron is less likely to segregate into clumps in the crystal and also has a lower diffusivity, where diffusivity is the ability of the dopant atoms to move in the crystal at elevated temperature. Pure undoped silicon has a room-temperature resistivity $> 10 \text{ k}\Omega\text{cm}$. Through doping, resistivities from this value all the way down to $< 1 \text{ m}\Omega\text{cm}$ can be realized, which represents more than seven orders of magnitude of tunability.

5.2 PN JUNCTION

A PN junction is created by putting a p-type doped semiconductor in contact with an n-type doped semiconductor. Most commonly, they are made from the same host material. In an unbiased PN junction, the chemical potentials on both sides of the junction are at the same energy, which means that the acceptor level of the p-type material lines up with the donor level of the n-type material; this is illustrated in [Figure 5.5](#).

When a voltage is applied across the PN junction, the side connected to the negative terminal of the voltage source is shifted up in energy by an amount q_0V . This represents adding an energy of q_0V to the negatively charged electrons on this side of the junction. (Since the zero of potential energy is arbitrary, this is the same as shifting the side connected to the positive terminal of the voltage source down in energy by q_0V .) In [Figure 5.6](#), we show a voltage source with the negative terminal

connected to the n-type side, resulting in increase in energy of q_0V relative to the p-type side. If $q_0V \ll E_g$, mobile electrons at the bottom of the conduction band on the p-type side face forbidden gap states on the n-type side, and likewise mobile holes at the top of the valence band on the n-type side face forbidden gap states on the p-type side. As a result, the mobile charges are trapped on their respective sides of the junction.

When q_0V reaches a value close to E_g , the conduction bands and the valence bands of the two materials become aligned. This means that, in response to the applied electric field, mobile electrons can freely flow from the conduction band in the n-type side to the conduction band in the p-type side, and mobile holes can freely flow from the valence band in the p-type side to the valence band in the n-type side, resulting in a current. This produces a current-voltage curve, or IV curve, like the one shown in Figure 5.7.

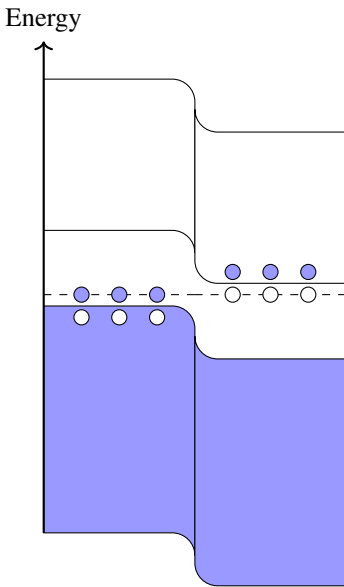


Figure 5.5 Energy diagram of an unbiased PN junction. The p-type side is on the left and the n-type side is on the right.

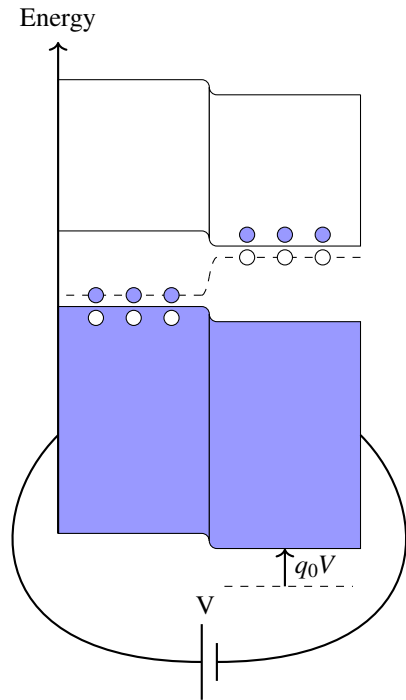


Figure 5.6 Energy diagram of a PN junction with bias voltage V applied across the junction.

When the voltage source has its positive terminal connected to the p-type side of the junction, this is known as forward-bias and corresponds to positive voltages in Figure 5.7. When the negative terminal is connected to the p-type side, this is known as reverse-bias and corresponds to negative voltages in Figure 5.7. In the forward-bias polarity, when $q_0V \approx E_g$, there is a rapid increase in the current. The voltage at

which this rapid increase in current occurs is called the threshold voltage V_T or the diode drop voltage. For silicon at room temperature, $E_g = 1.12$ eV and $V_T \approx 0.7$ V. If the voltage is increased beyond V_T , the current increases exponentially; this rise is so rapid that it is not possible to increase the voltage significantly beyond V_T because doing so would result in such a large current that the junction would burn out.

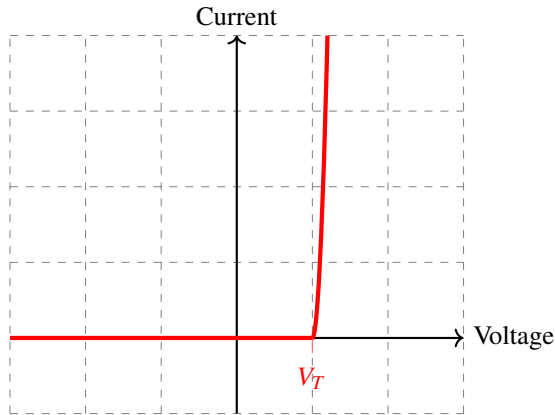


Figure 5.7 Current-voltage (IV) curve of a PN junction. Positive voltages correspond to the positive terminal of the voltage source being connected to the p-type side of the junction.

We can also consider a spatial representation of our doped semiconductors, which is shown in [Figure 5.8](#). In the p-type material, we have a random distribution of negative acceptor ions, which are immobile. We also have an equal number of positive valence-band holes, which are mobile. Likewise, in the n-type material, we have a random distribution of positive donor ions, which are immobile, along with an equal number of negative conduction-band electrons, which are mobile. When we place these two materials into contact, the mobile electrons and mobile holes near the interface will be attracted to each other and will recombine. This leaves the p-type side of the interface with a net negative charge and the n-type side of the interface with a net positive charge, producing an electric field across the interface, as shown in [Figure 5.9](#). The region where this electric field exists is known as the depletion region, as it is depleted of mobile charges. The electric field in the depletion region prevents further electron-hole recombination. This electric field corresponds to the energy barrier in our energy diagram of [Figure 5.5](#).

If we apply a voltage across our PN junction, the external voltage source creates its own electric field across the junction. The net electric field across the junction is then the vector sum of the internal depletion region electric field and the external electric field. If the electric field from the external voltage source points in the opposite direction as the depletion region electric field, this corresponds to forward-bias, as shown in [Figure 5.10](#). In this polarity, when the magnitude of the external electric field is equal to the magnitude of the internal electric field, the net electric field

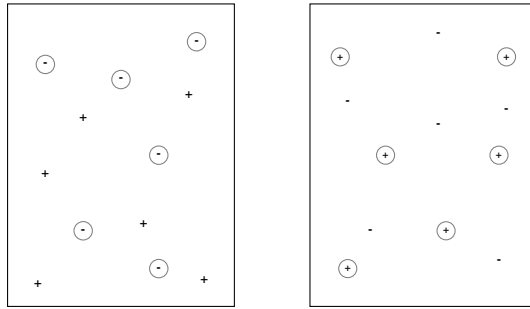


Figure 5.8 Spatial representation of p-type (left) and n-type (right) doped semiconductors. The $-$ and $+$ signs with circles around them represented immobile negative and positive ions, respectively, and the $-$ and $+$ charges without circles represent mobile conduction-band electrons and valence-band holes, respectively.

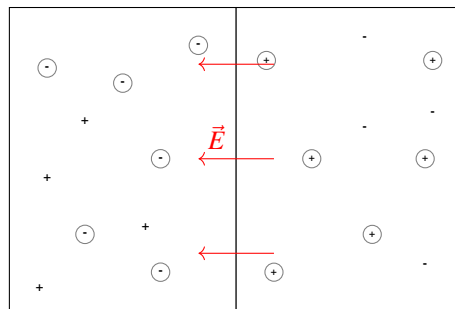


Figure 5.9 Spatial representation of an unbiased PN junction. Mobile electrons and holes near the interface recombine, resulting in a net electric field pointing from the n-type side of the interface to the p-type side. The region with the electric field is known as the depletion region, as it is depleted of mobile charges.

becomes zero and the depletion region disappears. Further increasing the voltage results in a rapid increase in current, as seen in the IV curve. If the external electric field points in the same direction as the internal electric field, this corresponds to reverse-bias, as shown in [Figure 5.11](#). Here the net electric field across the junction and hence the width of the depletion region grows larger compared to the unbiased case. This larger electric field corresponds to increasing the potential barrier in our energy diagram representation.

The PN junction is our first example of a *nonlinear* circuit component. It is nonlinear because the resistance of the component varies as a function of the voltage across the component. Any circuit component that produces an IV curve similar to the one shown in [Figure 5.7](#) is known as a *diode*. Today the vast majority of diodes are made from PN junctions, but there are other ways to produce such an IV curve; the very first diodes were made from vacuum tubes in the early 1900s.

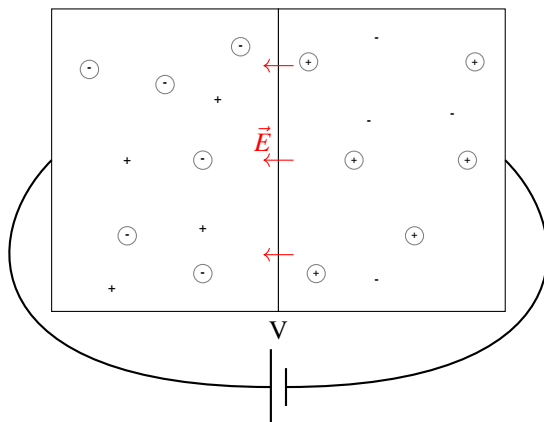


Figure 5.10 Spatial representation of forward-biased PN junction with $V < V_T$.

If we increase the voltage in the reverse-bias polarity, eventually we will reach what is known as the breakdown voltage. This is the voltage that results in dielectric breakdown at the junction interface. Generally one wants to avoid reaching the breakdown voltage, as doing so often destroys the diode. Most diodes have a breakdown voltage well above 20 V. A Zener diode is a special type of diode that is designed to make use of the breakdown voltage (without destroying the diode), and one can buy Zener diodes with different breakdown voltage values.

5.3 DIODE CIRCUITS

The circuit symbol for a diode is shown in Figure 5.12. The arrow indicates the direction that current flows when the diode is on. The side of the diode with the arrow tail is sometimes called the *anode* and side of the diode with the arrow head is sometimes called the *cathode*; these terms are a historical artifact from the days when diodes were made from vacuum tubes.

When the voltage across a diode is less than V_T (assuming a positive voltage corresponds to the forward-bias polarity), then almost no current flows through the diode, meaning that its resistance is extremely large. In this regime, we can often treat the diode as if it is an open circuit, i.e. as if it has infinite resistance. When the voltage across the diode reaches V_T , current will flow, and the voltage across the diode becomes “pinned” at approximately V_T ; it is not possible to increase the voltage significantly above V_T , and the diode resistance will be whatever is necessary to ensure that this is the case.

Our basic approach to analyzing diodes in circuits will be to consider how the circuit behaves in these two regimes – which can refer to as “diode on” and “diode off” for short – and then to determine under what condition the diode will switch from one regime to the other.

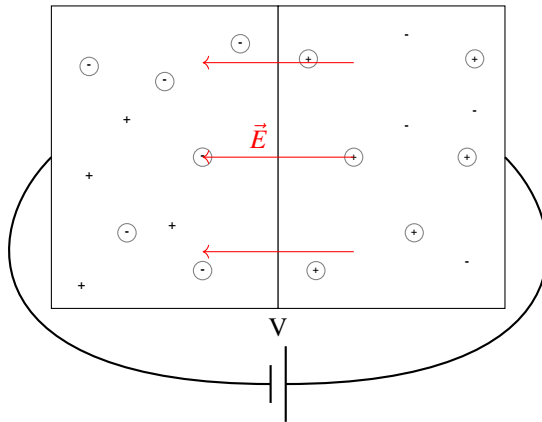


Figure 5.11 Spatial representation of reverse-biased PN junction.

As an example, consider the circuit shown in [Figure 5.13](#). In this circuit, we are assuming that V_{in} is an adjustable DC voltage source. When the diode is off, $R_{diode} \sim \infty$, and hence $V_{out} = V_{in}$. When V_{in} reaches V_T , the diode threshold voltage, the diode will turn on, and as V_{in} is increased further, V_{out} will remain “pinned” or “clamped” at V_T . From the loop rule, we know that whatever portion of the source voltage V_{in} is not across the diode must be across the resistor R , so with the diode on we can write the voltage across the resistor as $V_R = V_{in} - V_T$. A sketch of V_{out} as a function of V_{in} is shown in [Figure 5.14](#). This circuit is known as a voltage clamp because it prevents the output voltage from rising above V_T .

It is possible to shift the value at which V_{out} clamps by adding a DC voltage source in series with the diode, as shown in [Figure 5.15](#). In this circuit, $V_{out} = V_{in}$ when $V_{in} < (V_T + V_0)$ and $V_{out} = (V_T + V_0)$ when $V_{in} \geq (V_T + V_0)$.

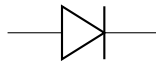


Figure 5.12 Circuit symbol for a diode. The arrow indicates the direction of possible current flow.

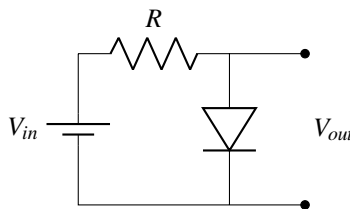


Figure 5.13 Voltage clamp circuit that prevents V_{out} from exceeding the diode threshold voltage V_T .

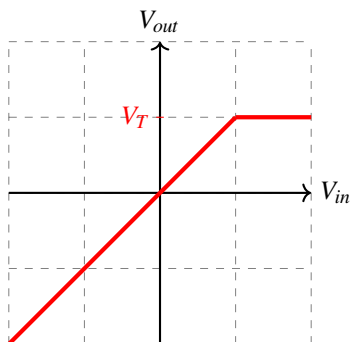


Figure 5.14 V_{out} as a function of V_{in} for voltage clamp circuit shown in Figure 5.13.

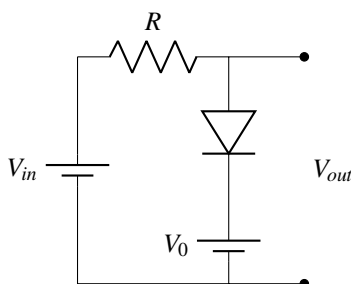


Figure 5.15 Voltage clamp circuit that prevents V_{out} from exceeding $(V_T + V_0)$.

Exercise 5.1 Finding the Diode Current

In the circuit shown in Figure 5.15, if $V_{in} = 3.5$ V, $V_0 = 1.0$ V, $R = 1.0$ k Ω , and the diode has a threshold voltage $V_T = 1.0$ V, what is the current through the diode?

Solution: Unless it is otherwise specified, we assume that V_{out} is measured with an ideal voltmeter. This means that the resistor and diode are in series and hence must have the same current. So if we find the current through the resistor, we have also found the current through the diode. Since $V_{in} > (V_T + V_0)$, we know that the diode is on. With the diode on, from the loop rule we know that $V_{in} - IR - V_T - V_0 = 0$. Solving for the current and plugging in the given values, we get $I = 1.5$ mA.

If we replace the DC voltage source in Figure 5.13 with an AC voltage source with a zero-to-peak amplitude of 2.0 V, this would result in the $V_{out}(t)$ shown in Figure 5.16. In this circuit, V_{out} follows V_{in} when $V_{in} < V_T$ and V_{out} “clamps” at V_T

when $V_{in} \geq V_T$. Here we assume that $V_T = 1.0$ V.

If we switch the diode and the resistor in the circuit of [Figure 5.13](#), we get the circuit shown in [Figure 5.17](#), which is known as a half-wave rectifier. By convention, a positive V_{in} means that the top side of the AC source is higher in potential than the bottom side. When the diode is off, all the source voltage is dropped across the diode, so $V_{out} = 0$. When the diode is on, $V_{out} = V_{in} - V_T$. The diode will be off whenever $V_{in}(t) < V_T$, and the diode turns on when $V_{in}(t)$ reaches V_T , resulting in the plot shown in [Figure 5.18](#), which assumes $V_T = 0.6$ V.

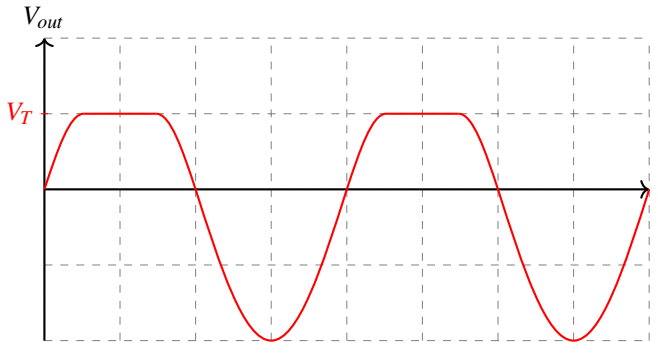


Figure 5.16 $V_{out}(t)$ for the voltage clamp circuit shown in [Figure 5.13](#) with the DC voltage source replaced by an AC voltage source with a zero-to-peak amplitude of 2 V.

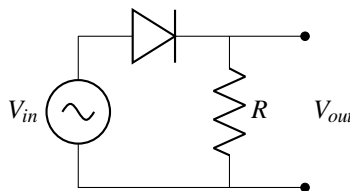


Figure 5.17 Half-wave rectifier circuit.

We see that the half-wave rectifier preserves the positive portion of the sinusoidal V_{in} (less V_T) and cuts off the negative portion. We can reverse this – i.e. preserve the negative portion (less V_T) and cut off the positive portion – by reversing the orientation of the diode in the circuit.

A full-wave rectifier preserves the positive portion of the sinusoidal V_{in} (less $2V_T$) and inverts the negative portion (also less $2V_T$); such a circuit is shown in [Figure 5.19](#). $V_{in}(t)$ and $V_{out}(t)$ for this circuit are shown in [Figure 5.20](#) assuming $V_T = 0.6$ V for all four diodes. When $V_{in}(t) > 2V_T$, diodes *A* and *D* will be on and diodes *B* and *C* will be off. This results in the simplified circuit shown in [Figure 5.21](#). Similarly, when $V_{in}(t) < -2V_T$, diodes *B* and *C* will be on and diodes *A* and *D* will be off, resulting

in the simplified circuit shown in Figure 5.22. When $|V_{in}(t)| < 2V_T$, all four diodes will be off and V_{out} will be zero.

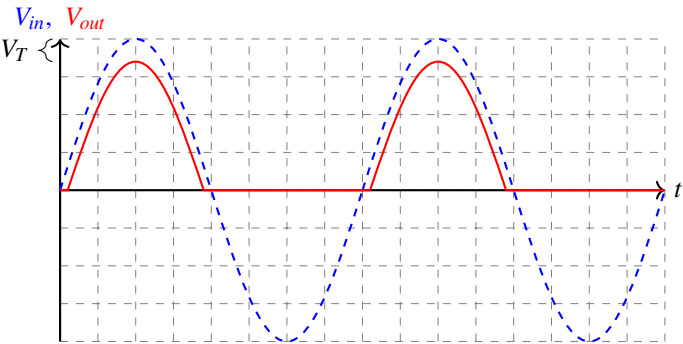


Figure 5.18 $V_{in}(t)$ (blue dashed line) and $V_{out}(t)$ (red solid line) for for half-wave rectifier circuit.

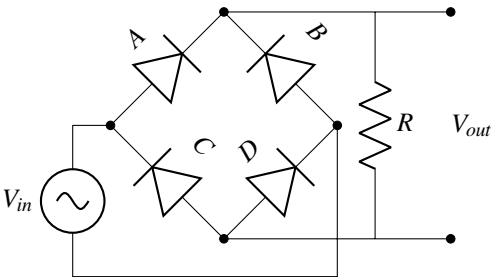


Figure 5.19 Full-wave rectifier circuit; connections at wire junctions are indicated with a dot.

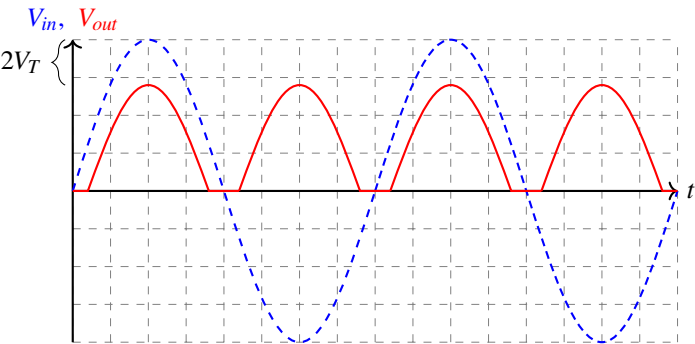


Figure 5.20 $V_{in}(t)$ (blue dashed line) and $V_{out}(t)$ (red solid line) for for full-wave rectifier circuit.

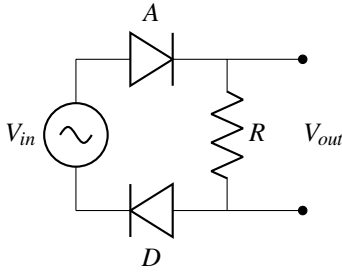


Figure 5.21 Simplified version of full-wave rectifier circuit when diodes *A* and *D* are on and diodes *B* and *C* are off (and are taken out of the schematic); this applies when $V_{in} > 2V_T$

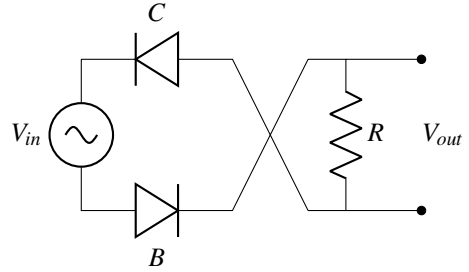


Figure 5.22 Simplified version of full-wave rectifier circuit when diodes *B* and *C* are on and diodes *A* and *D* are off (and are taken out of the schematic); this applies when $V_{in} < -2V_T$.

5.4 FREQUENCY MIXING

Near the threshold voltage V_T , the diode IV curve in Figure 5.7 can be approximated by the exponential function

$$I(V) = I_0 e^{\frac{q_0(V-V_T)}{k_B T}} \quad (5.3)$$

where I_0 is a constant, q_0 is the fundamental charge, k_B is Boltzmann's constant, and T is the temperature. The voltage-dependent term $e^{q_0 V / k_B T}$ can be expanded using a Taylor series of the form:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

The nonlinear terms ($n > 1$) in this expansion give rise to a phenomenon called *frequency mixing*. For simplicity, we will consider here just the quadratic ($n = 2$) term, although one can extend the following discussion to include higher-order terms as well.

If the voltage across the diode is the sum of two sources $V_1(t) = V_{10} \cos(\omega_1 t)$ and $V_2(t) = V_{20} \cos(\omega_2 t)$, then the quadratic term yields:

$$I(V) \propto (V_1(t) + V_2(t))^2 = V_1(t)^2 + V_2(t)^2 + 2V_1(t)V_2(t)$$

Using the product-to-sum identity, we can write the product term as:

$$\begin{aligned} 2V_1(t)V_2(t) &= 2V_{10}V_{20} \cos(\omega_1 t) \cos(\omega_2 t) \\ &= V_{10}V_{20} [\cos((\omega_1 - \omega_2)t) + \cos((\omega_1 + \omega_2)t)] \end{aligned}$$

We see that we get a term at the sum frequency $(\omega_1 + \omega_2)$ and a term at the difference frequency $(\omega_1 - \omega_2)$.

This result is in contrast to a linear circuit, whose response never contains frequency components other than the source frequencies. In general, a circuit or component with a nonlinear IV curve will produce sum and difference frequencies, with the amplitude of these frequency components depending on the exact shape of the IV curve.

If $\omega_1 = \omega_2$, then the difference frequency term will be at zero frequency (DC). In this case, we get a DC voltage that is proportional to the amplitude of our AC signal(s). This is known as *homodyne* mixing. Related to this is homodyne detection, in which we have a single input signal $V(t) = V_0 \cos(\omega t)$; the quadratic term in the IV curve will produce an output proportional to $V(t)^2$, and the resulting difference-frequency term from the product-to-sum identity will be a DC voltage that is proportional to the AC amplitude V_0 .

If $\omega_1 \neq \omega_2$, this is known as *heterodyne mixing*. One common use of heterodyne mixing is to down-convert a high-frequency signal to some lower frequency that is more convenient for analysis (e.g. a frequency that is low enough to run through an analog-to-digital converter for analysis on a computer). In this case, we want ω_1 and ω_2 to be close in frequency so that $|\omega_1 - \omega_2| \ll \omega_1, \omega_2$. In heterodyne mixing, it is common to call the high-frequency signal at ω_1 that one wants to down-convert the *RF* (for *radio frequency*), the signal at ω_2 the *LO* (for *local oscillator*), and the output signal at $|\omega_1 - \omega_2|$ the *IF* (for *intermediate frequency*). A circuit schematic for a heterodyne mixer (*mixer* for short) with the inputs and outputs labeled according to these conventions is shown in Figure 5.23.

One application of heterodyne mixers is in radio receivers. An AM radio signal consists of some radio frequency carrier signal that has been amplitude modulated by a lower frequency audio signal. This amplitude modulation can be accomplished by taking the product of the two signals. If we write the carrier as $V_C = V_{C0} \cos(\omega_C t)$ and the audio signal as $V_A = V_{A0} \cos(\omega_A t)$, then, using the product-to-sum identity, our product can be written as $V_C V_A = \frac{1}{2} V_{C0} V_{A0} [\cos((\omega_C - \omega_A)t) + \cos((\omega_C + \omega_A)t)]$. This signal is received by our radio. To demodulate this signal and recover the audio-frequency signal, our radio couples this into the *RF* port of a mixer along with an *LO* at a frequency ω_C , producing a difference frequency term at the *IF* output of the mixer at ω_A . Higher frequency terms at the output of the mixer can be eliminated with a low-pass filter.

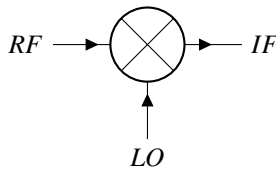


Figure 5.23 Circuit symbol for a heterodyne mixer that takes in an RF signal at frequency ω_1 and an LO signal at frequency ω_2 and produces an IF output at frequency $|\omega_1 - \omega_2|$.

When one adjusts the dial on a radio, one is changing the frequency of the *LO*; when the *LO* frequency matches the carrier frequency of a particular radio station, we are able to hear the demodulated audio signal from that station. In practice, most radios use two mixing steps rather than one; the reader who is curious to know more about this is encouraged to look up the term *superheterodyne receiver*.

5.5 SPECIAL TOPIC: THE LOCK-IN AMPLIFIER

A lock-in amplifier is a type of AC voltmeter that uses homodyne detection. Specifically, a lock-in amplifier takes the product of a signal $V_S = V_{0S} \cos(\omega t + \phi_S)$ and a reference at the same frequency $V_{ref} = V_{0ref} \cos(\omega t + \phi_{ref})$ and produces a difference-frequency output $V_{out} = (1/2)V_{0S}V_{0ref} \cos(\phi_S - \phi_{ref})$. The result is a DC voltage that is proportional to both the signal amplitude and the phase difference between the signal and the reference. The AC component of the output is removed with a low-pass filter. The device “locks in” on the reference frequency, which can be produced either by the lock-in itself or by an external source. A lock-in amplifier is known as a phase-sensitive detector because it not only rejects noise that is at different frequencies than the reference, it can also reject noise that is at the same frequency but out-of-phase. A high quality lock-in amplifier can measure AC voltages as small as ~ 10 nV.

A lock-in amplifier can also be used to measure a complex impedance. Most lock-in amplifiers have both an in-phase channel, which is proportional to $\cos(\phi_S - \phi_{ref})$, as well as an out-of-phase channel, which is proportional to $\cos(\phi_S - \phi_{ref} + \pi/2)$. If we take the reference voltage produced by the lock-in and place a very large resistor R_{bias} in series with this, it behaves as an AC current source whose current amplitude is equal to the reference voltage amplitude divided by R_{bias} . The AC current is then sent through some external component with impedance Z , and the resulting voltage across the external component at the reference frequency is measured by the lock-in amplifier. The reading on the in-phase channel will be proportional to the real part of Z and the reading on the out-of-phase channel will be proportional to the imaginary part of Z . Knowing the values of the reference frequency, the reference amplitude, and R_{bias} , we can then determine the value of the complex Z .

5.6 TRANSISTORS

A transistor is a three-terminal device. There are two general types of transistors, the bipolar junction transistor (BJT) and the field effect transistor (FET). In both cases, one terminal acts as a control knob to adjust the conductivity between the other two terminals. For a BJT, the three terminals are called the *base*, the *collector*, and the *emitter*, and adjusting the base current changes the conductivity between the collector and emitter. For a FET, the three terminals are called the *gate*, the *source*, and the *drain*, and adjusting the gate voltage changes the conductivity between the source and the drain.

There are many variants of FETs. These include the MOSFET (metal-oxide-semiconductor FET), the JFET (junction FET), and the FinFET (where *fin* refers

to the shape of the transistor). An example of an FET circuit symbol is shown in Figure 5.24.

There are two types of BJTs: the NPN BJT and the PNP BJT. The NPN BJT is made from putting a thin layer of p-type doped semiconductor in between two layers of n-type doped semiconductor. Likewise, the PNP BJT is made from putting a thin layer of n-type doped semiconductor in between two layers of p-type doped semiconductor. Circuit symbols for the NPN and PNP BJTs are shown in Figures 5.25 and 5.26. For the sake of brevity, we will focus our discussion on just one type of transistor, the NPN BJT, and then we will consider how this transistor can be used to build several types of circuits.

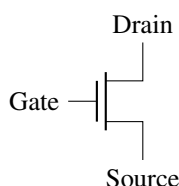


Figure 5.24 Example of a circuit symbol for a field effect transistor

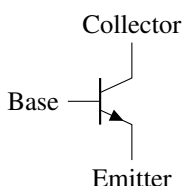


Figure 5.25 Circuit symbol for a NPN bipolar junction transistor

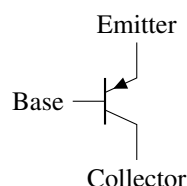


Figure 5.26 Circuit symbol for a PNP bipolar junction transistor.

5.6.1 NPN BIPOLAR JUNCTION TRANSISTOR

To understand the operation of an NPN BJT, we will use the same spatial representation that we used in our discussion of the PN junction. This spatial representation of the NPN BJT is shown in Figure 5.27. This figure shows the transistor connected to an external circuit in what is known as the common emitter configuration. To operate, the transistor's base-emitter junction must be forward-biased at the threshold voltage V_T and the collector-base junction must be reverse-biased.

Electrons from the emitter are *injected* into the base through the forward-biased base-emitter junction. (Remember that the direction of the current is opposite the direction in which electrons move.) In the base, mobile electrons will *diffuse*, and those that reach the depletion region in the vicinity of the collector-base junction will be *swept* by the electric field into the collector. To increase the probability that electrons will reach the depletion region in the base, the p-type base should be lightly doped and relatively thin. The electrons that are swept into the collector then flow out of the collector into the external circuit. In this way, we have electrons that flow from the emitter through the base to the collector, and hence a current from the collector through the base to the emitter.

Increasing V_{CE} pushes the collector-base junction further into the reverse-bias regime, growing the size of the depletion region. The farther the depletion region extends into the base, the greater the probability that electrons diffusing in the base will reach the depletion region and be swept into the collector. Thus we expect to see an increase in I_C as V_{CE} is increased. As V_{CE} continues to be increased, at some point

the depletion region will extend nearly the full width of the base, resulting in a nearly 100% probability that mobile electrons in the base will reach the depletion region. When this happens, we do not expect I_C to increase significantly as V_{CE} is increased further. This results in one of the IV curves seen in Figure 5.28, in which we plot I_C on the y-axis and V_{CE} on the x-axis. The region in which I_C increases significantly with increasing V_{CE} is called the *saturation region*, while the region in which I_C is relatively constant with increasing V_{CE} is known as the *active region*.

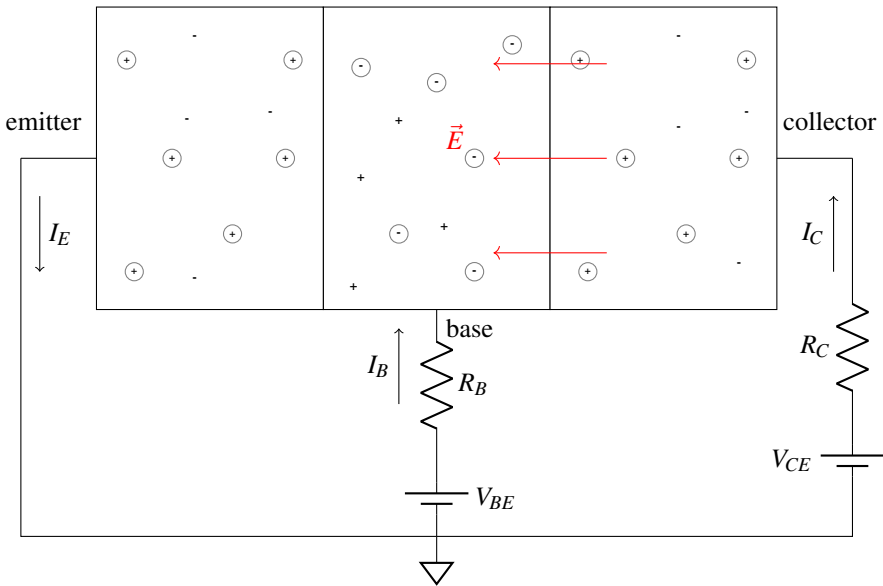


Figure 5.27 Spatial representation of a NPN BJT connected in the common emitter configuration. For proper operation, the base-emitter junction must be forward-biased by V_{BE} and the collector-base junction must be reverse-biased by V_{CE} .

If we increase I_B by increasing V_{BE} , this pushes the base-emitter junction further into forward-bias, increasing the rate at which electrons are injected from the emitter into the base. This causes I_C to increase across all values of V_{CE} . Conversely, decreasing I_B causes I_C to decrease for all values of V_{CE} . This produces a family of IV curves like those shown in Figure 5.28, with each individual curve corresponding to a particular value of I_B . The numerical values in the figure are intended as representative examples, but the actual values, as well as the exact shape of the IV curve, will vary from transistor to transistor.

All types of transistors produce IV curves that are qualitatively similar to what we see in Figure 5.28. They all have a saturation region and an active region, and the curves shift up or down as the “control knob” is changed; for the BJT, the control knob is the base current, while for the FET the control knob is the gate voltage.

Two important device applications become apparent from these IV curves. The first is that this device can be used as an electrical switch. If one is biased in the

active region, changing I_B can effectively turn the collector-emitter channel on (for higher I_B) and off (for I_B near zero). We also see that, in the active region, relatively small changes in I_B produce much larger changes in I_C . This effect can be used to make an amplifier; if a small AC signal is coupled to I_B , this will produce much larger oscillations at the same frequency in I_C .

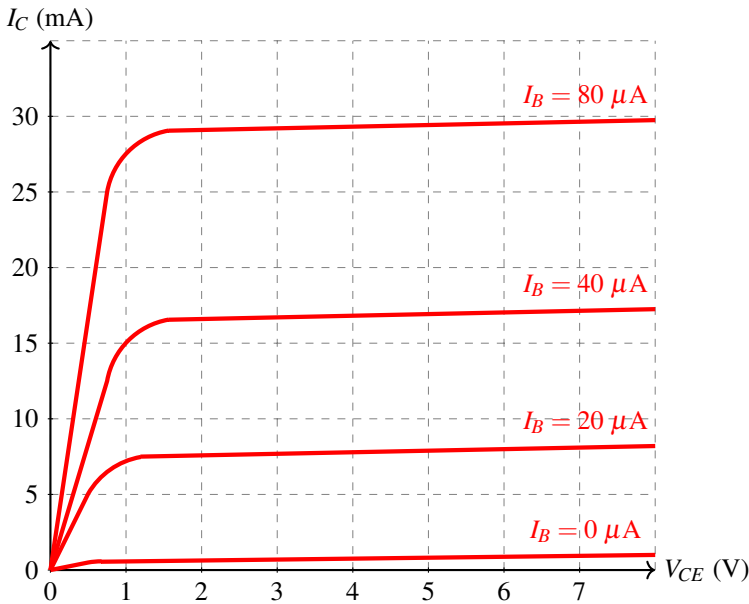


Figure 5.28 Representative current-voltage (IV) curves of NPN BJT.

5.6.2 TRANSISTOR CIRCUITS

For the bipolar junction transistor, we define the unitless parameters $\beta = I_C/I_B$ and $\alpha = I_C/I_E$. The value of β is typically in the range of 20–200, while the value of α is typically in the range of 0.95–0.995. β is known as the *current gain*, and it is sometimes represented by the symbol h_{FE} (the *FE* is for *forward emitter*). From the junction rule, we know that $I_C + I_B = I_E$, i.e. the current flowing into the transistor must equal the current flowing out of the transistor. This can be combined with our expressions for α and β to get $\beta = \alpha/(1 - \alpha)$ and $\alpha = \beta/(1 + \beta)$. The values of β for different values of I_C are often specified on a transistor's datasheet.

A simple transistor circuit is shown in [Figure 5.29](#). If the voltage across the forward-biased base-emitter junction is 1.0 V, then the emitter voltage is $V_1 - 1$. Hence the current through resistor R , which is also the emitter current, is $I_E = (V_1 - 1)/R$. The collector current is $I_C = \alpha I_E$, so if we assume that $\alpha \approx 1$, then I_C , which is also the current through R_L , is $I_C \approx (V_1 - 1)/R$. In other words, the value of the current through R_L is independent of the value of R_L , and this circuit acts as a constant current source through R_L with the value of the current determined by V_1

and R . This behavior will hold as long as the transistor remains properly biased, i.e. as long as the base-emitter junction is forward-biased and the collector-base junction is reverse-biased.

Figure 5.30 shows a transistor circuit known as an emitter follower. Here we will consider what happens if we couple to the base a DC voltage plus an AC voltage. The DC component of the base voltage V_B will result in a DC component of the emitter voltage V_E that is equal to V_B minus the threshold voltage of the forward-biased base-emitter junction. The AC component of the base voltage, which we will denote ΔV_B , will produce an equivalent AC voltage at the emitter, i.e. $\Delta V_E = \Delta V_B$. If we consider ΔV_B as the input signal and ΔV_E as the output signal, then our circuit has a transfer function $H = 1$ for AC signals. This may seem rather useless, but the utility of this circuit becomes more apparent when we consider the input impedance seen at the base and the output impedance seen at the emitter.

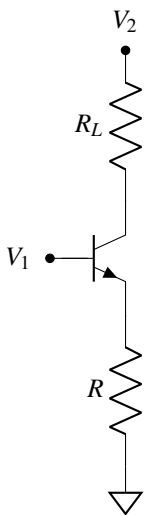


Figure 5.29 Transistor current source circuit.

To find the input impedance Z_{in} at the base, we will initially assume that the output at the emitter is connected to an ideal (infinite resistance) voltmeter. The AC current at the output is then $\Delta I_E = \Delta V_E / R_E$. Using the fact that $\Delta V_E = \Delta V_B$, we can write $\Delta I_E = \Delta V_B / R_E$. We then use the definition of β , $\beta \Delta I_B = \Delta I_C$, and the junction rule relation $\Delta I_C + \Delta I_B = \Delta I_E$ to write $\Delta I_E = \Delta I_B (1 + \beta)$. Combining this with our previous expression for ΔI_E gives us $\Delta I_E = \Delta V_B / R_E = \Delta I_B (1 + \beta)$, which can be rearranged as:

$$\frac{\Delta V_B}{\Delta I_B} = R_E (1 + \beta)$$

Of course $\Delta V_B / \Delta I_B$ is just the input impedance Z_{in} . Because β is large, we see that the effect of the circuit is to create a large input impedance by multiplying the value of R_E by a factor $(\beta + 1) \approx \beta \approx 100$. If the output is measured not by an ideal

voltmeter but by a voltmeter with internal resistance R_M , then our expression for the input impedance becomes:

$$Z_{in} = (1 + \beta) \left(\frac{1}{R_E} + \frac{1}{R_M} \right)^{-1}$$

To find the output impedance Z_{out} , we can determine the Thevenin equivalent resistance measured between the emitter and ground with the input connected to a voltage source with internal resistance R_S . In this case, the Thevenin equivalent AC voltage $\Delta V_{TH} = \Delta V_E = \Delta V_B$. The short-circuit AC current is $\Delta I_S = \Delta I_E = \Delta I_B (1 + \beta) = (\Delta V_B / R_S) (1 + \beta)$. Then $Z_{out} = \Delta V_{TH} / \Delta I_S = R_S / (1 + \beta)$. We see that the output impedance is equal to the source resistance divided by a factor of $(1 + \beta)$. The large value of β helps to ensure that Z_{out} is a small value, even for a relatively nonideal voltage source.

If we want to connect the output of circuit *A* to the input of circuit *B* and the output impedance of *A* is not much smaller than the input impedance of *B*, then we will not get maximum voltage transfer. Placing an emitter follower circuit in between *A* and *B* will serve to increase the voltage transfer by effectively scaling down the output impedance of *A* and scaling up the input impedance of *B* by a factor of $(1 + \beta)$.

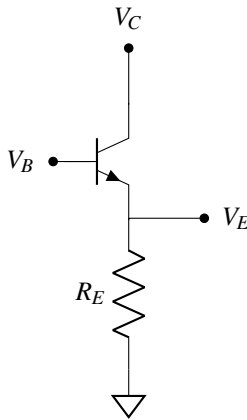


Figure 5.30 Emitter follower circuit.

Exercise 5.2 Darlington Pair

The two-transistor circuit in [Figure 5.31](#) is known as a Darlington pair. Show that this circuit behaves like a single transistor with the base *B*, collector *C*, and emitter *E* as labeled that has a β value that is approximately equal to the product of the β values of the two individual transistors.

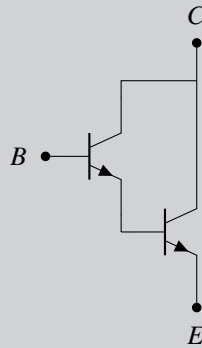
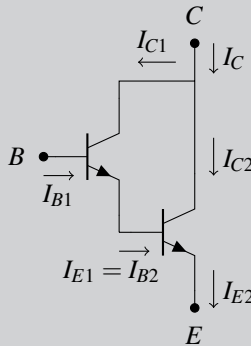


Figure 5.31 Darlington pair.

Solution: We begin by labeling the currents as shown:



For the left transistor we have $\beta_1 = I_{C1}/I_{B1}$ and for the right transistor we have $\beta_2 = I_{C2}/I_{B2}$. For the total circuit, we have $\beta = I_C/I_{B1}$. Applying the junction rule to each transistor individually and to the whole circuit, we can write the following equations:

$$I_{B1} + I_{C1} = I_{E1}$$

$$I_{B2} + I_{C2} = I_{E2}$$

$$I_{B1} + I_C = I_{E2}$$

Using the fact that $I_{E1} = I_{B2}$, we can combine the first and second junction rule equations to write $I_{B1} + I_{C1} = I_{E2} - I_{C2}$. Using our expressions for β_1 and β_2 gives $I_{B1} + \beta_1 I_{B1} = I_{E2} - \beta_2 I_{B2}$. We can solve this for I_{E2} and then equate it with the third junction rule expression to get $I_C = \beta_1 I_{B1} + \beta_2 I_{B2}$. Plugging in $I_C = \beta I_{B1}$ and simplifying, we get $\beta = \beta_1 + \beta_2 (I_{B2}/I_{B1})$. We can then substitute $I_{B2} = I_{E1} = I_{B1} + I_{C1}$ to get

$$\beta = \beta_1 + \beta_2 (1 + I_{C1}/I_{B1}) = \beta_1 + \beta_2 + \beta_1 \beta_2$$

Since $\beta_1, \beta_2 \gg 1$, this can be approximated as

$$\beta \approx \beta_1 \beta_2$$

A transistor can function as an amplifier when it is biased in the active region of the transistor's IV curve (Figure 5.28). In this region, relatively small changes in the base current produce much larger changes in the collector current. A type of transistor amplifier known as a common-emitter amplifier is shown in Figure 5.32. In this circuit, V_{CC} is a positive DC voltage that biases the transistor in its active region. The resistors R_1 and R_2 function as a voltage divider, resulting in a DC base voltage that is approximately equal to $V_{CC}R_2/(R_1 + R_2)$.

The two capacitors are known as coupling capacitors; they are designed to block DC signals but pass AC signals. The coupling capacitors ensure that the DC biasing condition does not depend on what is connected to V_{in} and V_{out} . The input voltage V_{in} and the output voltage V_{out} must be AC signals, as they must pass through the coupling capacitors.

Like we saw with our emitter follower circuit, the AC component of the emitter voltage is equal to the AC component of the base voltage, $\Delta V_E = \Delta V_B$, where ΔV_B is the same as our AC input signal provided that the voltage drop across the coupling capacitor is negligible. Then we can write

$$\frac{\Delta V_B}{R_E} = \frac{\Delta V_E}{R_E} = \Delta I_E$$

If we assume that $\alpha \approx 1$, then $\Delta I_E \approx \Delta I_C$, and so we can write

$$\Delta I_C \approx \Delta V_B / R_E$$

We also know that the AC component of the collector voltage is $\Delta V_C = -\Delta I_C R_C$. Assuming the voltage drop across the coupling capacitor is negligible, this is also our AC output voltage. The negative sign comes from the fact that a larger current through R_C corresponds to a smaller voltage at V_C relative to V_{CC} . Combining this expression for ΔV_C with our previous expression for ΔI_C , we get

$$\frac{\Delta V_B}{R_E} = -\frac{\Delta V_C}{R_C}$$

This can be rearranged to solve for $\Delta V_C / \Delta V_B$, which is also the ratio of the AC output voltage to the AC input voltage, i.e. our transfer function, which is

$$H = \frac{-R_C}{R_E}$$

The negative sign in our expression for H means that this is an *inverting* amplifier, i.e. the output voltage will be π phase-shifted relative to the input voltage. The amplitude of the AC output voltage will be a factor of R_C / R_E larger than the amplitude of the AC input voltage.

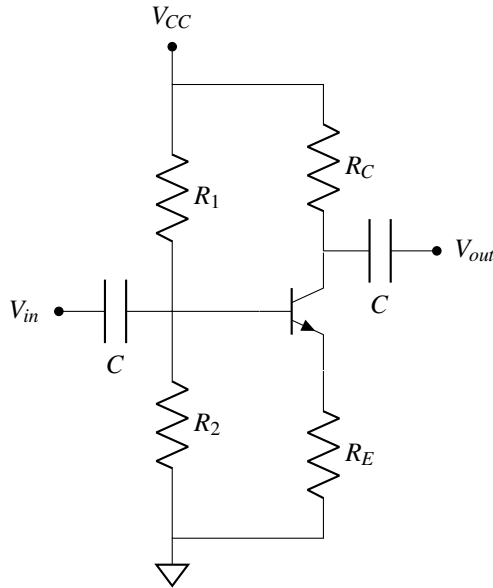


Figure 5.32 Common emitter amplifier circuit.

All amplifiers have a maximum output voltage that they can produce, an effect known as *saturation*. In our common emitter amplifier, the output is the AC component of the collector voltage. In order for the transistor to operate, the collector voltage must be higher than the base voltage, whose DC component is approximately $V_{CC}R_2/(R_1 + R_2)$. The upper limit of the collector voltage is set by V_{CC} , as we must have $V_C < V_{CC}$ in order to have a current flowing into the collector. These two limits define the maximum possible oscillation amplitude of V_C and hence the maximum possible amplitude of our output signal, which is known as our saturation value. The saturation value can be increased by increasing V_{CC} , although increasing V_{CC} increases the power dissipation, and if we push V_{CC} too high we risk burning out the transistor.

The differential amplifier circuit shown in [Figure 5.33](#) contains two transistors labeled Q_1 and Q_2 . It takes input voltages V_1 and V_2 and produces the output voltage V_{out} . It is called a differential amplifier because V_{out} is proportional to the difference between V_1 and V_2 . One reason that differential amplifiers can be useful is that any noise that is common to both inputs will subtract out at the output. V_{CC} is a positive DC voltage that biases the two transistors in their active region and also determines the saturation value of the amplifier.

We assume that the two transistors are identical and that each has a fixed voltage V_T across its forward-biased base-emitter junction. This means that the emitter voltage of transistor Q_1 is $V_1 - V_T$ and the emitter voltage of transistor Q_2 is $V_2 - V_T$. We call the emitter current from transistor Q_1 I_1 and the emitter current from transistor Q_2 I_2 . Assuming the base current is negligible, then the collector current and emitter

currents are equal. Denoting the voltage at the junction where resistor R_1 connects to the two R_E resistors as V_J , we can use Ohm's law to write $I_1 = (V_1 - V_T - V_J)/R_E$ and $I_2 = (V_2 - V_T - V_J)/R_E$. We can plug these expressions for I_1 and I_2 into the expression for the current flowing through resistor R_1 , which is $I_1 + I_2 = (V_J + V_{CC})/R_1$, to get:

$$\frac{V_1 - V_T - V_J}{R_E} + \frac{V_2 - V_T - V_J}{R_E} = \frac{V_J + V_{CC}}{R_1}$$

The collector current through Q_2 can be written as $I_2 = (V_{CC} - V_{out})/R_E$. Combining this with our previous expression for I_2 gives:

$$\frac{V_2 - V_T - V_J}{R_E} = \frac{V_{CC} - V_{out}}{R_E}$$

We now have two equations with two unknowns, V_J and V_{out} . Solving both equations for V_J , equating them, and then solving for V_{out} yields:

$$V_{out} = \frac{R_1 [R_C(V_1 - V_2) + 2R_E V_{CC}] + R_E [R_E V_{CC} - R_C(V_2 + V_{CC} - V_T)]}{R_E(R_E + 2R_1)}$$

If we assume that $R_1 \gg R_E$, then $(R_E + 2R_1) \approx 2R_1$ and we can write:

$$V_{out} \approx \frac{R_C}{2R_E}(V_1 - V_2) + V_{CC} \left(1 + \frac{R_E}{2R_1} - \frac{R_C}{2R_1} \right) + \frac{R_C}{2R_1}(V_T - V_2)$$

Since $R_1 \gg R_E$, $R_E/(2R_1) \approx 0$. If we choose $R_C = 2R_1$, then the term with V_{CC} becomes approximately zero, leaving us with:

$$V_{out} \approx \frac{R_C}{2R_E}(V_1 - V_2) + (V_T - V_2)$$

Since $R_C = 2R_1$ and $R_1 \gg R_E$, it must be that $R_C \gg R_E$, which means that the first term will generally be large compared to the second term. This gives us our final simplified expression for the output voltage of our differential amplifier:

$$V_{out} \approx \frac{R_C}{2R_E}(V_1 - V_2)$$

5.7 SPECIAL TOPIC: INTEGRATED CIRCUITS

An integrated circuit is a circuit that is entirely constructed on a single chip, typically a semiconductor chip. The integrated circuit was first developed by Jack Kilby and Robert Noyce in the late 1950s, and the technology has steadily advanced in the years since then. This advance is the basis of Moore's law, a prediction made by Gordon Moore in 1965 that the number of transistors on integrated circuit chips would double every one to two years. The development of integrated circuits has been approximately consistent with Moore's prediction over the past half-century, and this development has been central to the modern electronics industry.

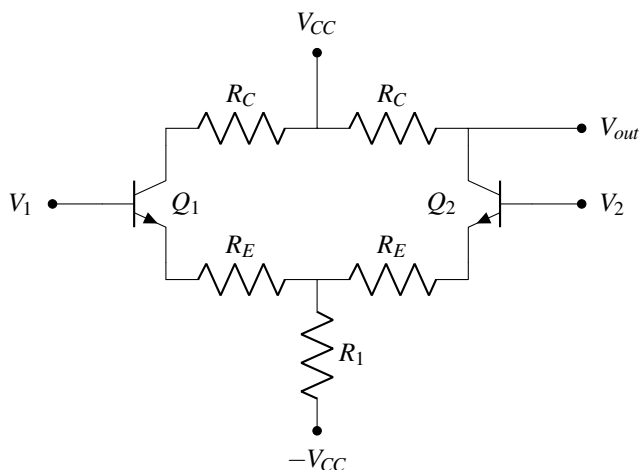


Figure 5.33 Differential amplifier circuit.

Today's computer processor chips can have upwards of 100 million transistors per square millimeter of chip area. In order to have so many transistors in such a small area, the transistors must have nanometer-scale dimensions. The process of creating integrated circuits is known as microfabrication.

Microfabrication involves coating the chip with a thin layer of a polymer known as photoresist, or resist for short. Ultraviolet (UV) light chemically alters the resist. The resist-coated chip is exposed to UV light through a patterned mask. The chip is then placed in a chemical developer that removes either the resist that was exposed to the UV light (for positive resists) or the resist that was not exposed (for negative resists). In this way, the mask pattern is transferred to the resist. The patterned resist layer is then used for a processing step such as depositing a thin film of material, removing exposed material via a process called etching, or doping of the surface of the exposed semiconductor via ion implantation. After the processing step, the remaining resist is chemically removed. Through a sequence of multiple such patterning and processing steps, a complex integrated circuit can be built.

The fabrication of integrated circuits is typically performed on 300 mm (≈ 12 inch) diameter silicon wafers. These wafers are cut from large single crystals known as boules and polished to achieve a surface roughness of less than 1 nm. A 12 inch silicon wafer that has been patterned with microfabricated circuits is shown in [Figure 5.34](#). After the integrated circuit fabrication, the wafer is then diced (cut) into many individual chips, with each chip known as a die. A single die is shown in [Figure 5.35](#). Each die is then enclosed inside a standardized package that allows it to be mounted onto a circuit board. There are a number of different types of packages; one example, also shown in [Figure 5.35](#), is an 8-pin dual inline package (DIP).

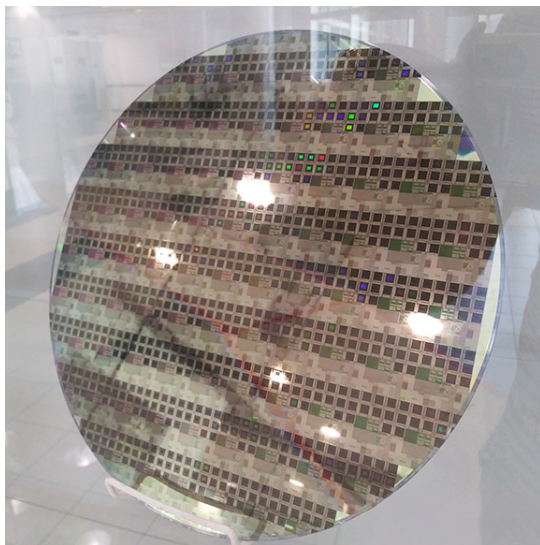


Figure 5.34 12 inch diameter silicon wafer with microfabricated integrated circuits.

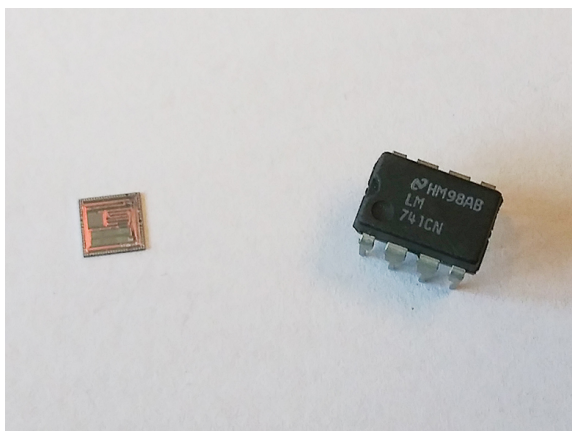
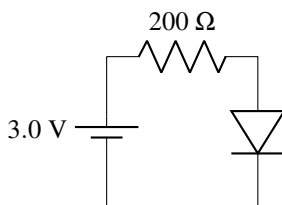


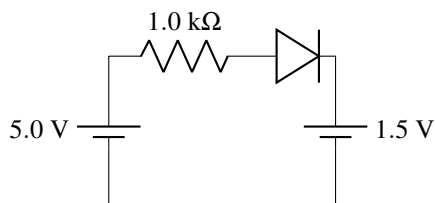
Figure 5.35 Approximately 4 mm $\times$ 4 mm integrated circuit die (left) and integrated circuit die packaged in 8-pin dual in-line package for mounting on a circuit board (right).

5.8 PROBLEMS

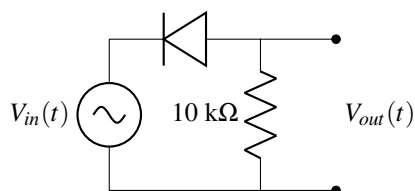
Problem 1. In the circuit shown below, find (a) the voltage across the 200 Ω resistor and (b) the current through the diode. Assume the diode threshold voltage is 1.0 V.



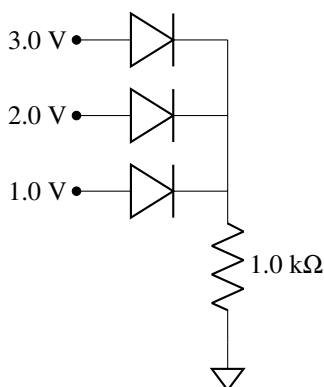
Problem 2. In the circuit shown below, find the current through the diode. Assume the diode threshold voltage is 0.70 V.



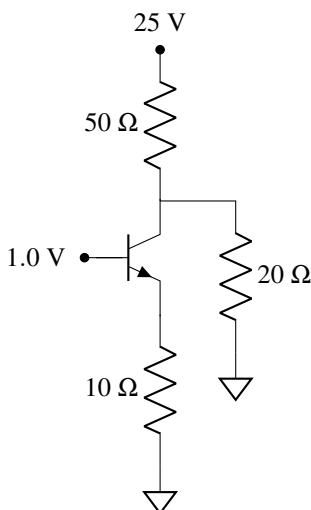
Problem 3. In the circuit shown, V_{in} is an ideal AC voltage source with a peak-to-peak amplitude of 4.0 V, and the diode has a threshold voltage of 0.5 V. Sketch V_{in} and V_{out} as functions of time on the same plot; show two full periods.



Problem 4. In the circuit shown, find the value of the current through the $1.0\text{ k}\Omega$ resistor. Assume each diode has a threshold voltage of 0.5 V. The points labeled 1.0 V, 2.0 V, and 3.0 V are each being held at the specified potential (defined with respect to the circuit ground) by a DC voltage source. (Hint: It may not be obvious at first glance which diodes are on and which are off. You may want to make a guess, solve based on this guess, and then check your results for validity. If your results are not valid, try a different guess.)



Problem 5. For the NPN BJT circuit shown, find the value of the current through the $20\ \Omega$ resistor. Assume that the voltage across the forward-biased base-emitter junction is 0.7 V and that the emitter current is approximately equal to the collector current (i.e. assume $\alpha = 1$).



Problem 6. Consider an *NPN* transistor for which the collector voltage is held at 12 V , the base voltage is held at 1.5 V , the emitter is connected to ground through a $110\ \Omega$ resistor, and $\beta = 100$. If the voltage across the forward-biased base-emitter junction is 0.6 V , find the values of the base current I_B , the collector current I_C , and the emitter current I_E .



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

6 Operational Amplifiers

The operational amplifier, or op-amp, is a differential voltage amplifier with an extremely large transfer function magnitude, a.k.a. voltage gain, typically $> 10^5$. This is known as the open-loop gain, and it is designed to be so large that it is quasi-infinite such that any difference in voltage between the two inputs will drive the output into saturation. Op-amps are designed to have a very large input impedance – to approximate an ideal voltmeter – and a very small output impedance – to approximate an ideal voltage source.

As we saw in [chapter 5](#), a differential amplifier can be built from transistors. Indeed, op-amps are usually built from transistors, resistors, and possibly a capacitor or two. When using op-amps, however, we usually treat them as a discrete component and do not worry about what is inside the op-amp; we assume that the manufacturer designed the op-amp properly and it behaves as described in the datasheet provided by the manufacturer. While we could build our own op-amp, a general purpose op-amp costs less than a dollar, so it is not likely to be worth our time to do so.

In treating the op-amp as a discrete component, we are going up a level in the circuit hierarchy. The bottom-most level in the circuit hierarchy is composed of fundamental components, including linear components such as resistors, inductors, and capacitors, as well as nonlinear components such as diodes and transistors. The next level up are devices that are built from these fundamental components, which includes op-amps. When designing more complex circuits, it is often advantageous to work at the highest level of the circuit hierarchy that one can; building complex circuits from only fundamental components can be an unnecessarily complicated and labor-intensive process.

The op-amp is an active device, meaning that it must be connected to a power supply in order to operate. It is powered by a positive DC supply voltage $+V_0$ and a negative DC supply voltage $-V_0$. These supply voltages serve to bias the transistors in the op-amp in the active region of their IV curve. Typical supply voltages are in the range of ± 8 to ± 18 V. The op-amp's saturation voltage – the largest positive and negative voltages that the output can produce – are equal to or a bit smaller than the supply voltages. (The supply voltages are sometimes called *rails*, and the case when the saturation voltages are equal to the supply voltages is known as rail-to-rail operation.)

A circuit symbol for an op-amp is shown in [Figure 6.1](#). V_+ is called the noninverting input, V_- is called the inverting input, and V_{out} is the output. With no feedback, we have

$$V_{out} = A(V_+ - V_-) \quad (6.1)$$

where A is the open-loop voltage gain. If $V_+ > V_-$, then we expect V_{out} to be in positive saturation, and if $V_- > V_+$, then we expect V_{out} to be in negative saturation.

The DC supply voltages $\pm V_0$ are shown in the left-hand version but are omitted in the right-hand version. In circuit schematics, it is common to omit the op-amp supply voltage connections for simplicity, but we still assume that the op-amp is provided with appropriate supply voltages.

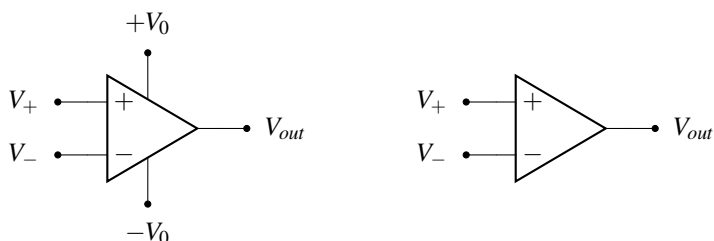


Figure 6.1 Op-amp circuit symbol with (left) and without (right) the supply voltage connections shown.

6.1 IDEAL OP-AMPS WITH STABLE FEEDBACK

Feedback is a general concept in which the output of a system affects its input. In an op-amp circuit, this means that there must be some external electrical connection between the op-amp output and at least one of its inputs. There are two types of feedback: stabilizing and destabilizing. If the system is perturbed from equilibrium, stabilizing feedback works to bring the system back to equilibrium, while destabilizing feedback pushes the system further away from equilibrium.

In op-amp circuits, stabilizing feedback requires an electrical connection between the op-amp output and the inverting input. We can understand why this is the case by considering a voltage fluctuation at the output. If it is a positive fluctuation, feedback will result in an increase in the voltage at the inverting input V_- , which, from Equation 6.1, will cause V_{out} to decrease, bringing it back toward equilibrium. Conversely, if the feedback connects V_{out} to the noninverting input V_+ , then a positive fluctuation in V_{out} causes V_+ to increase, which in turn causes V_{out} to increase further, resulting in a runaway effect that ends with V_{out} in positive saturation.

The purpose of stabilizing feedback is to reduce the circuit gain from the open-loop gain A to some more reasonable value that is determined by external components. In this way, we don't need to worry about the exact value of A , provided that it is sufficiently large, and we can set the voltage gain to be whatever value we want through the use of a small number of external components. For analyzing op-amp circuits with stabilizing feedback, we often make two assumptions that simplify the process of analyzing the circuit. These assumptions are the following:

- (1) V_{out} will be whatever is necessary to ensure that $V_- = V_+$.
- (2) No current flows into the op-amp inputs.

Assumption 1 is valid for stabilizing feedback as long as the necessary value of V_{out} is smaller than the saturation voltage. Assumption 2 corresponds to treating the

op-amp inputs as ideal voltmeters. In a real op-amp, the input impedance is large but not infinite, so in reality some current does flow into the inputs. But the value of that current is sufficiently small, typically in the range of 1 – 100 nA, that it can often be treated as approximately zero. Now we will use these two assumptions to analyze several op-amp circuits. In the next section we will consider what happens when these assumptions are no longer valid.

The first circuit we will consider is shown in Figure 6.2 and is a noninverting amplifier. Assumption 2 states that no current flows into the inputs, which means that V_{out} and V_- are related by the voltage divider formula: $V_- = V_{out}R_2/(R_1 + R_2)$. Assumption 1 means that $V_- = V_+ = V_{in}$. Hence we have $V_{in} = V_{out}R_2/(R_1 + R_2)$. Rearranging this, we get

$$\frac{V_{out}}{V_{in}} = \frac{R_1}{R_2} + 1 \quad (6.2)$$

Since this is a four-terminal circuit, we can identify the ratio of V_{out} to V_{in} as the transfer function H .

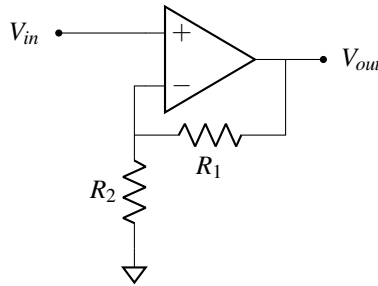


Figure 6.2 Noninverting amplifier circuit.

In Figure 6.3, we see the circuit for an inverting amplifier. Because of assumption 1, $V_- = V_+ = 0$. We can then use Ohm's law to write expressions for the currents through R_1 and R_2 which, because of assumption 2, must be equal, and so we equate these two expressions:

$$\frac{V_{in}}{R_1} = \frac{-V_{out}}{R_2}$$

Here we have assumed that the current flows through each resistor from left to right, although if we had assumed the other direction for the current flow, that would just correspond to multiplying both sides of the equation by -1 . Rearranging this to solve for $V_{out}/V_{in} = H$, we get

$$H = \frac{-R_2}{R_1} \quad (6.3)$$

Figure 6.4 shows an adder circuit. We can write Ohm's law expressions for the currents through the two resistors labeled R_1 , which, because of assumption 2, must be the same current, and hence we have $(V_1 - V_+)/R_1 = (V_+ - V_2)/R_1$. We see that V_{out} is related to V_- by our familiar voltage divider formula: $V_- = V_{out}R_3/(R_2 + R_3)$.

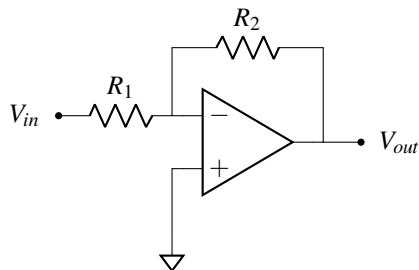


Figure 6.3 Inverting amplifier circuit.

Using the fact that $V_- = V_+$ (assumption 1), we can combine these two equations and solve for V_{out} , which gives us:

$$V_{out} = \frac{1}{2} (V_1 + V_2) \left(\frac{R_3 + R_2}{R_3} \right) \quad (6.4)$$

Since this is not a four-terminal circuit, we cannot define a transfer function in the usual way. This circuit has two inputs – V_1 and V_2 (both defined with respect to the common circuit ground) – and produces an output V_{out} that is proportional to their sum. While the value of the resistors R_1 do not appear in our expression for V_{out} , the presence of those two resistors is still necessary, as they allow V_1 and V_2 to be different voltages.

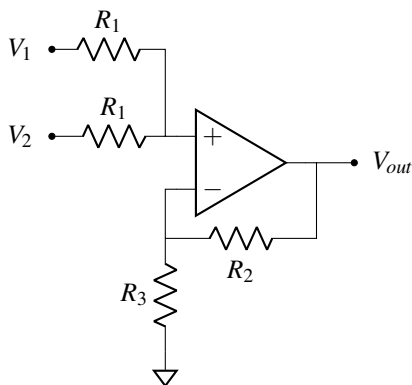


Figure 6.4 Adder circuit.

So far all of our op-amp circuits have used only op-amps and resistors, resulting in frequency-independent behavior (at least in the ideal case). Of course, we can also build op-amp circuits using components that have a frequency-dependent impedance such as inductors and capacitors. As an example, let's consider the inverting amplifier of [Figure 6.3](#) with a capacitor placed in parallel with resistor R_2 ; this is shown in [Figure 6.5](#).

We can analyze this circuit using the same approach as we did for the inverting amplifier except with R_2 replaced by an equivalent impedance from the parallel combination of resistor R_2 and capacitor C , i.e. $Z_{eq} = 1/(1/R_2 + i\omega C)$. The resulting transfer function is

$$H = \frac{-Z_{eq}}{R_1} = \frac{-R_2}{R_1} \frac{1}{1 + i\omega R_2 C} \quad (6.5)$$

In the limit that $\omega \rightarrow 0$, this becomes $H = -R_2/R_1$, i.e. the inverting amplifier result that we found without the capacitor. In the limit that $\omega \rightarrow \infty$, $H \rightarrow 0$. This circuit is known as an active low-pass filter; it amplifies signals in the low-frequency pass-band and attenuates signals in the high-frequency stop-band. One can also use op-amps to create active versions of high-pass, band-pass, and band-stop filters.

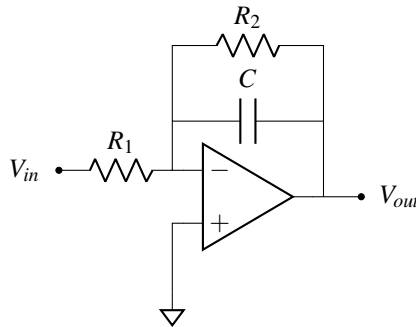


Figure 6.5 Active low-pass filter circuit.

6.2 NONIDEALITIES IN REAL OP-AMPS

The behavior of real op-amps can deviate from our ideal assumptions, and we should be mindful of these deviations when designing circuits. Here we will discuss some of the most significant nonidealities in real op-amps, but this discussion is not exhaustive.

The maximum rate at which the output voltage of an op-amp can change is given by the op-amp's *slew rate*, which has units of volts per second (V/s). The slew rate, along with the output signal amplitude, determine the maximum frequency at which our op-amp circuit can operate; this is known as the gain-bandwidth product. As an example, consider an op-amp circuit that outputs a triangle wave with a period of 1 μ s. If the triangle wave has a peak-to-peak amplitude of 1 V, then the rising slope of the triangle wave is 2 V/ μ s (the peak-to-peak amplitude divided by half a period), and we would want an op-amp with a slew rate of at least 2 V/ μ s to avoid distorting the output waveform. If, however, the peak-to-peak amplitude is only 0.1 V, then we would only need a slew rate of at least 0.2 V/ μ s to avoid distorting the output waveform.

The *input offset voltage* is the DC voltage offset seen at the output of the op-amp circuit referred to its input. For the case of an amplifier circuit, this means dividing

the output offset voltage by the circuit's transfer function to refer it to the input. A straightforward way to check for this offset voltage is to ground the input of the op-amp circuit and measure the output; any voltage seen at the output in this case is from the input offset voltage. Most op-amp chips have two pins denoted as *offset null* or *offset trim*. One can use these pins to create a circuit to tune out any offset voltage. To do this, one connects each of the fixed connections of a potentiometer (often $\sim 10\text{ k}\Omega$) to these two pins, and then connects the wiper terminal of the potentiometer to a DC voltage source (often the positive or negative supply voltage). Adjusting the potentiometer then adjusts the offset voltage.

Our second assumption of ideal op-amp behavior was that no current flows into the op-amp inputs. In practice, some current does flow into the inputs; this current is known as the *input bias current*. The value of this current varies for different types of op-amps but is usually $< 100\text{ nA}$. If there is some equivalent resistance between an op-amp input and ground, then the input bias current will produce a voltage at the input that is equal to the product of the input bias current and the equivalent resistance; this will produce a corresponding offset voltage at the op-amp output. The op-amp's two inputs – the inverting input and the noninverting input – typically draw similar input bias currents, but the values are not necessarily exactly equal. The difference between the input bias currents drawn by the two inputs is known as the *input offset current*.

Commercial op-amps come with a datasheet that specifies typical values for these nonideal quantities (as well as others not discussed here). As an example, the LM741 is a general-purpose, inexpensive op-amp. Its datasheet specifies the following typical values: a slew rate of $0.5\text{ V}/\mu\text{s}$, an input offset voltage of 1.0 mV , an input bias current of 80 nA , and an input offset current of 20 nA . One can buy op-amps with higher slew rates and/or lower offsets, but these generally come with an increase in price.

6.3 OP-AMPS WITHOUT STABLE FEEDBACK

Without stable feedback, the op-amp output will always be in either positive or negative saturation. This behavior can be used to build several types of useful circuits. One type is known as a comparator.

A comparator compares an input voltage V_{in} to a reference voltage V_{ref} . If $V_{in} > V_{ref}$, it produces one output, and if $V_{in} < V_{ref}$, it produces a different output. We can make a simple comparator out of an op-amp by applying V_{in} to one input and V_{ref} to the other input. If V_{in} is applied to the noninverting input and V_{ref} is applied to the inverting input, then V_{out} will be in positive saturation ($+V_{sat}$) if $V_{in} > V_{ref}$ and V_{out} will be in negative saturation ($-V_{sat}$) if $V_{in} < V_{ref}$. Conversely, the outputs will be switched if we apply V_{in} to the inverting input and V_{ref} to the noninverting input; this case is known as an inverting comparator and is shown in Figure 6.6. A plot of a time-dependent V_{in} and a time-independent V_{ref} is shown in Figure 6.7 along with the resulting comparator output V_{out} . The rate at which the output signal switches between positive and negative saturation is set by the op-amp's slew rate.

If V_{in} is noisy, this can produce switching noise, as shown in Figure 6.8. This arises when the noise causes V_{in} to jump back and forth across V_{ref} when the two values are very close. One technique for eliminating switching noise is to introduce *hysteresis* in the value of V_{ref} . Hysteresis is a general phenomenon in which the state of a system has some dependence on its history. In this case, we mean that the value of V_{ref} will be different depending on whether V_{out} is in positive or negative saturation. This can be achieved through the introduction of nonstabilizing feedback. This nonstabilizing feedback is realized through a connection between the op-amp output and its noninverting input, as shown in Figure 6.9; this type of comparator circuit is known as a Schmitt trigger.

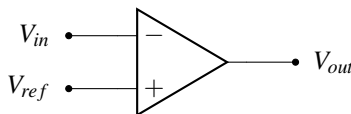


Figure 6.6 Inverting comparator circuit.

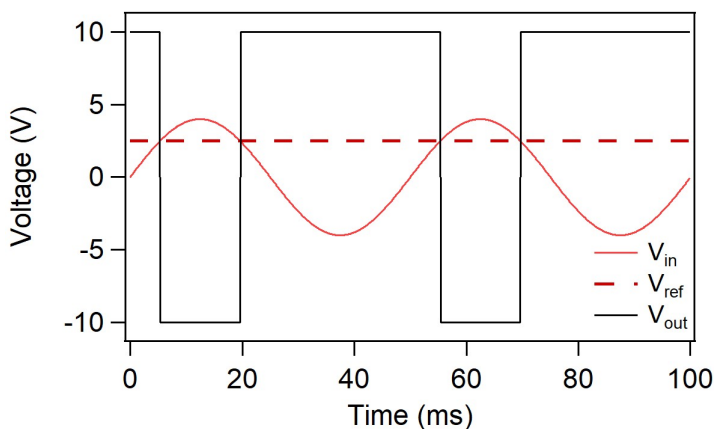


Figure 6.7 For the input and reference voltages shown in red, the inverting comparator produces the output shown in black. We have assumed that the op-amp saturates at ± 10 V and the slew rate is much faster than the time scale shown

In the Schmitt trigger circuit of Figure 6.9, V_S is a user-defined DC voltage, but we do not call it V_{ref} , as we want to keep the convention of V_{ref} being the voltage at the noninverting input. In this case, the value of V_{ref} is determined by the values of both V_S and V_{out} , and it is through this dependence on V_{out} that we realize hysteresis.

Given V_S and V_{out} , we can solve for V_{ref} under the assumption that the op-amp inputs draw no current. In this case, we can draw the feedback portion of the circuit as shown in Figure 6.10. We will get two values of V_{ref} , one for the case of V_{out} in positive saturation and one for the case of V_{out} in negative saturation. The difference between these two values of V_{ref} is the hysteresis. To avoid switching noise, one

wants the difference to be larger than the noise amplitude in V_{in} .

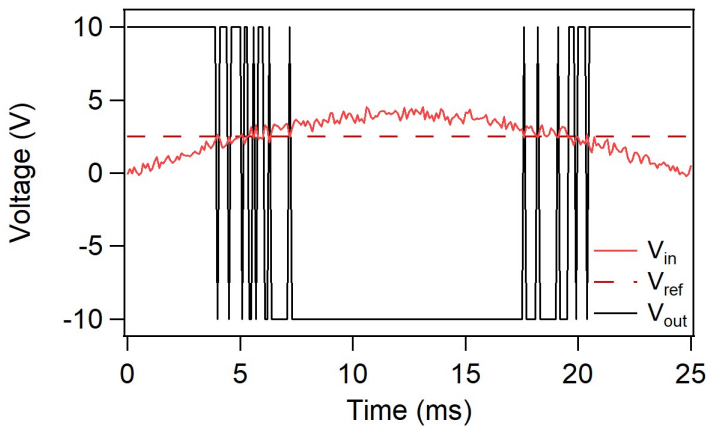


Figure 6.8 Example of the output of an inverting comparator showing switching noise. This is obtained by adding noise to the oscillatory V_{in} signal shown in [Figure 6.7](#) (note the change in the time scale between these two figures).

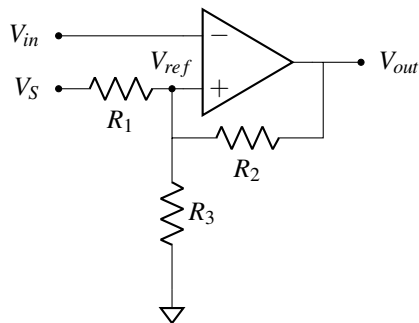


Figure 6.9 Schmitt trigger circuit.

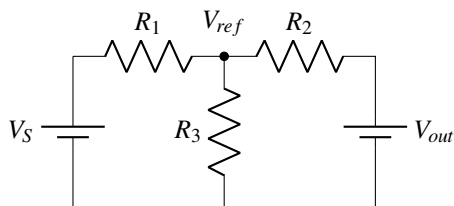
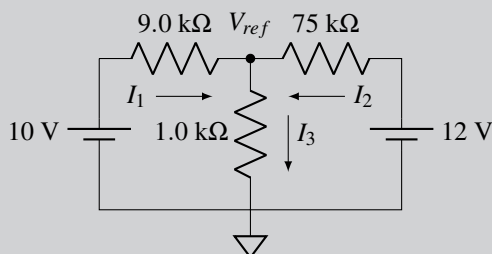


Figure 6.10 Feedback circuit from Schmitt trigger shown in [Figure 6.9](#).

Exercise 6.1 Schmitt Trigger Hysteresis

For the Schmitt trigger circuit in Figure 6.9, whose feedback circuit is shown in Figure 6.10, we have $R_1 = 9.0 \text{ k}\Omega$, $R_2 = 75 \text{ k}\Omega$, $R_3 = 1.0 \text{ k}\Omega$, and $V_S = 10 \text{ V}$. If the op-amp output saturates at $\pm 12 \text{ V}$, find the two values of V_{ref} .

Solution: We'll start with the positive saturation case, i.e. $V_{out} = +12 \text{ V}$. We will define the negative terminals of our two voltage sources as ground (0 V), the current through R_1 as I_1 , the current through R_2 as I_2 , and the current through R_3 as I_3 :



Then we can use Ohm's law to write an equation for the potential drop across each resistor:

$$10 - V_{ref} = 9,000I_1$$

$$12 - V_{ref} = 75,000I_2$$

$$V_{ref} = 1,000I_3$$

And, from the junction rule, we have:

$$I_1 + I_2 = I_3$$

We now have four independent equations and four unknowns, so we can proceed to solve via substitution. One way to do this is to solve each of the first three equations for the given current, plug each of these into the junction rule equation, and then solve the resulting equation for V_{ref} . When we do this, we get $V_{ref} = 1.13 \text{ V}$.

Next we need to solve for the negative saturation case. To do this, we can take our previous set of four equations and just replace 12 V with -12 V in the second equation. Making this change and solving, we get $V_{ref} = 0.846 \text{ V}$.

Note that if the $75 \text{ k}\Omega$ were removed, the two other resistors would form a divide-by-ten voltage divider, meaning that V_{ref} would be 1.0 V . The feedback realized by the connection between V_{out} and V_{ref} through the $75 \text{ k}\Omega$ resistor causes V_{ref} to increase slightly above 1.0 V when V_{out} is in positive saturation and to decrease slightly below 1.0 V when V_{out} is in negative saturation. This

means that, if V_{out} is initially in positive saturation, the output will switch when V_{in} exceeds the higher V_{ref} value of 1.13 V. But as soon as V_{out} switches, V_{ref} changes to its lower value of 0.846 V. As a result, fluctuations in V_{in} will not produce switching noise as long as these fluctuations are smaller than the difference in the V_{ref} values, which in this case is 0.28 V. This difference can be made larger or smaller by adjusting the value of the feedback resistor R_2 .

Next we will take the Schmitt trigger circuit from Figure 6.9 and remove V_S , V_{in} , and R_1 . Then we will add a feedback resistor R_f between V_{out} and the inverting input and a capacitor C between the inverting input and ground. The result, which is shown in Figure 6.11, is known as a *relaxation oscillator*. This circuit is inherently unstable, and it uses this instability to produce an oscillatory output signal.

The voltage at the noninverting input V_+ is related to V_{out} by the voltage divider equation: $V_+ = V_{out}R_3/(R_2 + R_3)$. At DC, the capacitor is an open circuit, so in this limit the voltage at the inverting input V_- should be equal to V_{out} . If V_{out} is in positive saturation, then the previous expressions give us $V_- > V_+$ (assuming R_2 is nonzero), which means that the output should be in negative saturation. Conversely, if V_{out} is in negative saturation, then we have $V_- < V_+$, which means that the output should be in positive saturation. This result seems to contain a fundamental contradiction.

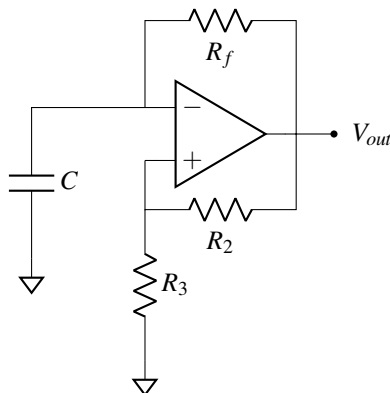


Figure 6.11 Relaxation oscillator circuit.

We can understand the circuit – and resolve the apparent contradiction – if we consider its time-dependent behavior. In this case, the capacitor is no longer just an open circuit. Let's assume that initially the op-amp is off and the voltage everywhere is zero. Then at time $t = 0$ we turn the op-amp on and the output becomes $V_{out} = -V_{sat}$, i.e. negative saturation. (We assume here that V_{sat} is a positive number.) This immediately makes $V_+ = V_{sat}R_3/(R_2 + R_3)$. The value of V_- does not change instantaneously; it will decrease from its initial value of zero to a value of $-V_{sat}$ in the limit $t \rightarrow \infty$, but the rate at which it changes is given by an exponential decay

with a time constant $\tau = R_f C$. The output will remain in negative saturation ($-V_{sat}$) until V_- drops below V_+ , at which point the output will switch to positive saturation ($+V_{sat}$). When this happens, V_+ will immediately switch to $V_+ = V_{sat}R_3/(R_2 + R_3)$, and V_- will begin to increase with the same exponential behavior. When V_- increases beyond the value of V_+ , the output will switch back to positive saturation. This periodic switching behavior will continue indefinitely, producing an oscillatory output.

We can find an expression for the period T of this oscillation. To do this, we'll call the moment that V_{out} switches from $+V_{sat}$ to $-V_{sat}$ time $t = 0$. Then, at $t = 0$, V_+ will be $-V_{sat}R_3/(R_2 + R_3)$ and V_- will still be at its previous value of $+V_{sat}R_3/(R_2 + R_3)$. As we move forward in time, V_- will decrease according to the exponential function:

$$V_-(t) = V_{sat} \left[\left(\frac{R_2 + 2R_3}{R_2 + R_3} \right) e^{\frac{-t}{R_f C}} - 1 \right]$$

The reader can check that this expression gives the appropriate results of $V_- = +V_{sat}R_3/(R_2 + R_3)$ at $t = 0$ and $V_- = -V_{sat}$ in the limit $t \rightarrow \infty$. Of course, V_- will never reach $-V_{sat}$ because once V_- reaches $V_+ = -V_{sat}R_3/(R_2 + R_3)$, the output will switch again. The time it takes V_- to decay from its value at $t = 0$ to the value of $V_+ = -V_{sat}R_3/(R_2 + R_3)$ is half a period, so we can equate $V_-(t = T/2) = V_+$, which gives us:

$$V_{sat} \left[\left(\frac{R_2 + 2R_3}{R_2 + R_3} \right) e^{\frac{-T/2}{R_f C}} - 1 \right] = -V_{sat}R_3/(R_2 + R_3)$$

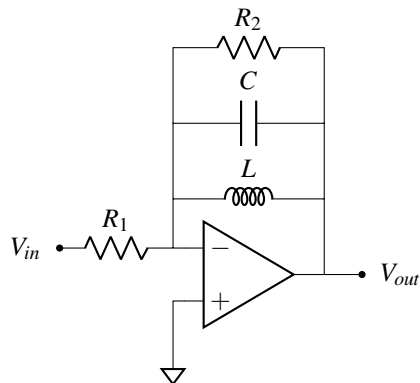
Solving this for the period T , we get:

$$T = 2R_f C \ln \left(\frac{R_2 + 2R_3}{R_2} \right)$$

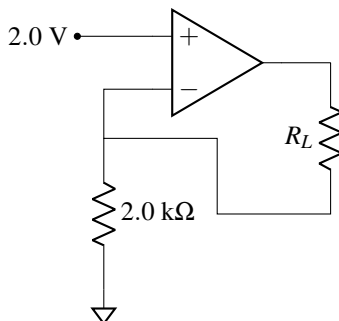
If we assume that the product of the oscillation amplitude and the oscillation frequency $f = 1/T$ is very small compared to the op-amp's slew rate, then $V_{out}(t)$ will be a square wave. We can set the frequency of this square wave via our choices of the component values C , R_f , R_2 , and R_3 .

6.4 PROBLEMS

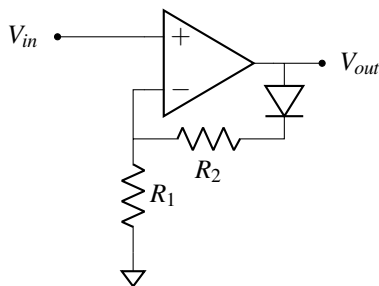
Problem 1. Assuming ideal op-amp behavior with stable feedback, find an expression for the magnitude of the transfer function $|H|$ in terms of the given circuit parameters. Be sure to fully simplify your answer.



Problem 2. (a) Assuming ideal op-amp behavior with stable feedback, what is the value of the current through R_L ? (Even though the value of R_L is unknown, you should be able to get a value for the current.) (b) If the op-amp output saturates at ± 10 V, what is the range of values of R_L for which the result from part (a) will be valid?

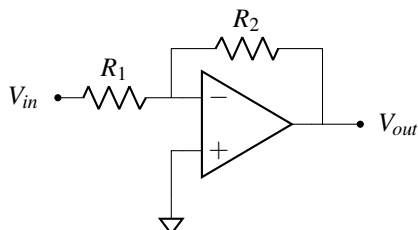


Problem 3. In the circuit shown, the diode has a threshold voltage of 1.0 V. Assuming ideal op-amp behavior with stable feedback, find an expression for V_{out} in terms of V_{in} , R_1 , and R_2 .

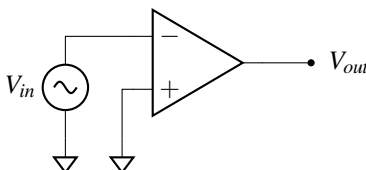


Problem 4. Using the rules for ideal op-amp behavior with stable feedback, the circuit shown is an inverting amplifier with an output voltage $V_{out} = -(R_2/R_1)V_{in}$.

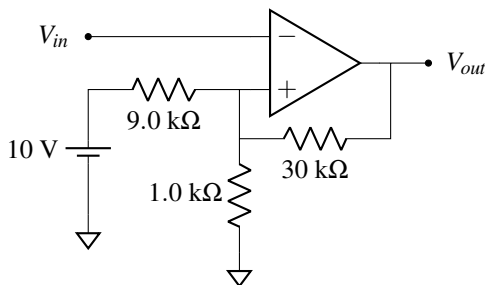
(a) If an input bias current I_b flows into each input of the op-amp, how does this modify the expression for V_{out} ? (Your new expression should include I_b .) (b) We can compensate for the effect of the input bias current by adding a resistor R_3 between the noninverting input and ground; this makes the voltage at the noninverting input $V_+ = -I_b R_3$. What value of R_3 will give us back our original expression for the output voltage $V_{out} = -(R_2/R_1)V_{in}$ that we found assuming ideal op-amp behavior? (You should get an expression for R_3 that is only in terms of R_1 and R_2 .)



Problem 5. A 2.0 V peak-to-peak sine wave is applied to the inverting input of an op-amp without feedback. Sketch both $V_{in}(t)$ and $V_{out}(t)$ for a time t corresponding to two full periods of the input signal. Assume that the op-amp saturates at ± 10 V and that the frequency of $V_{in}(t)$ is extremely small compared to the op-amp's slew rate.



Problem 6. For the Schmitt trigger circuit shown, find the value of V_{in} that will cause the output to switch if the output is (a) initially in positive saturation and (b) initially in negative saturation. Assume that the op-amp output saturates at ± 10 V.





Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

7 Noise

A central challenge of experimental science is extracting a weak signal from a noisy system. In this chapter we will consider both intrinsic and extrinsic sources of noise in electrical systems and some techniques for maximizing the signal-to-noise ratio (S/N) of our measurements.

The S/N of a measurement is proportional to the square root of the averaging time τ provided that:

- (1) the signal amplitude and the noise spectral density are constant during τ ,
and
- (2) the signal and the noise are uncorrelated.

For the case of a discrete number of measurements N rather than a continuous measurement, the S/N is proportional to the square root of N . In either case, this square root dependence means that reducing noise has significant benefits in terms of the time it takes to perform a measurement. For example, reducing the noise by a factor of 10 reduces the required measurement time to achieve a particular S/N by a factor of 100. If conditions **1** and **2** are not true, it will generally take a longer time to achieve a particular S/N .

7.1 QUANTIFYING NOISE

Consider a DC measurement of resistance. In general, one applies a known current I , measures the resulting voltage V , and determines the resistance from Ohm's law, $R = V/I$. If we look at the measured voltage V on a sufficiently sensitive voltmeter, we see that its value is not constant – it fluctuates in time.

The DC voltage V_{DC} is the time-average of the instantaneous voltage $V(t)$, $V_{DC} = \langle V(t) \rangle$, where $\langle \dots \rangle$ denotes a time average of the quantity inside the brackets. At each point in time we can calculate the deviation from the average $\delta V(t) = V(t) - \langle V(t) \rangle$. We can then quantify the fluctuations of a data set such as [Figure 7.1](#) by calculating $\langle \delta V^2 \rangle$. It is important to square δV before taking the time average; for a random noise source, δV is positive half the time and negative half the time, in which case $\langle \delta V \rangle = 0$. Those who are familiar with statistics may have noted that $\langle \delta V^2 \rangle$ is the variance of $V(t)$ and $\sqrt{\langle \delta V^2 \rangle}$ is the standard deviation of $V(t)$. We commonly assume that noise, e.g. $\delta V(t)$, can be modeled as a Gaussian-distributed random variable; while this is true for many sources of noise, there are exceptions. Even if the distribution is not Gaussian, it will often approach Gaussian behavior when the sample size is large; this is a phenomenon known as the central limit theorem.

The voltage noise spectral density S_V is defined as

$$S_V = \frac{\langle \delta V^2 \rangle}{\Delta f} \quad (7.1)$$

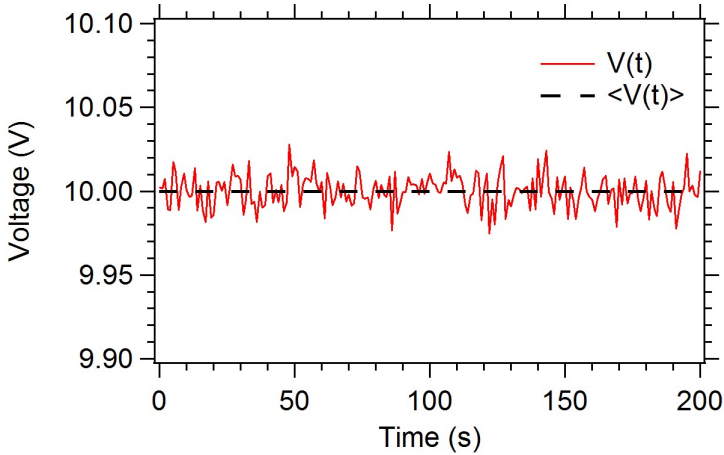


Figure 7.1 A sensitive measurement of a DC voltage signal as a function of time showing the noise.

where Δf is the bandwidth over which $\langle \delta V^2 \rangle$ is measured. S_V has units of V^2/Hz and is a measure of the average fluctuation amplitude squared per unit measurement bandwidth.

We can quantify current fluctuations in the same manner as voltage fluctuations, with $\delta I(t) = I(t) - \langle I(t) \rangle$ and the current noise spectral density $S_I = \langle \delta I^2 \rangle / \Delta f$.

A noise source is described as “white” if the noise is uniformly distributed across all frequencies, which means that $\langle \delta V^2 \rangle$ is linearly proportional to Δf . No noise source is truly uniform over *all* frequencies, but many noise sources are uniform across enough of a frequency range that we can approximate them as “white” in certain contexts.

7.2 JOHNSON-NYQUIST NOISE

All resistors exhibit voltage fluctuations that are proportional to their temperature. This is due to the random thermal motion of charge carriers and the resulting fluctuating electric fields. This phenomenon was first measured by John B. Johnson at Bell Labs in 1927 and was first described theoretically by Harry Nyquist at Bell Labs in 1928, and hence it is known as Johnson-Nyquist noise.

At low frequencies – defined as $hf \ll k_B T$, where f is the frequency, T is the temperature, h is Planck’s constant, and k_B is Boltzmann’s constant – the voltage spectral density of Johnson-Nyquist noise for a resistor with resistance R is given by:

$$S_{V,JN} = 4k_B T R \quad (7.2)$$

In this low-frequency limit, Johnson-Nyquist noise is frequency-independent or “white”. (At room temperature, $k_B T / h \approx 6$ THz, so we are safely in the low-frequency limit at microwave frequencies and below.) For a known resistance R , a

measurement of S_V tells us the temperature of the resistor. This is a useful technique for the determining the temperature of very small circuit elements and is known as Johnson noise thermometry.

Let's consider how we might measure Johnson-Nyquist noise. We can model our noisy resistor as a noise-free resistor R in series with a source of voltage fluctuations $\delta V_{JN}(t)$ that behaves as an ideal (i.e. zero impedance) voltage source whose output is a Gaussian-distributed random variable. We then connect this model of our noisy resistor to a measuring circuit that we model as an ideal voltmeter in parallel with some input resistance R_{in} . For now, we will neglect the Johnson-Nyquist noise of R_{in} , i.e. we will assume that the measuring circuit is noiseless. This circuit is shown in Figure 7.2.

This circuit is a voltage divider, so the measured fluctuations δV_{meas} are

$$\delta V_{meas} = \delta V_{JN} \frac{R_{in}}{R + R_{in}}$$

and hence

$$\langle \delta V_{meas}^2 \rangle = \langle \delta V_{JN}^2 \rangle \left(\frac{R_{in}}{R + R_{in}} \right)^2$$

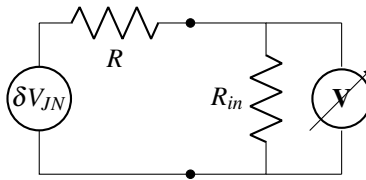


Figure 7.2 Model of a resistor R exhibiting Johnson-Nyquist noise (left) connected to a voltmeter with input impedance R_{in} (right).

First we'll consider the case where $R_{in} \rightarrow \infty$ (or, more realistically, $R_{in} \gg R$). This achieves maximum voltage transfer and is common in lower frequency circuit applications. In this case, we get $\langle \delta V_{meas}^2 \rangle = \langle \delta V_{JN}^2 \rangle = 4k_B T R \Delta f$.

Next we'll consider the case where $R_{in} = R$. This achieves maximum power transfer and is common in microwave-frequency applications. In this case, we get $\langle \delta V_{meas}^2 \rangle = \langle \delta V_{JN}^2 \rangle / 4 = k_B T R \Delta f$. If we measure the power P_N over a bandwidth Δf , we get $P_N = \langle \delta V_{meas}^2 \rangle / R = k_B T \Delta f$.

Sometimes the noise generated by an electrical component is quantified as a noise temperature T_N , defined as

$$T_N = \frac{P_N}{k_B \Delta f} \quad (7.3)$$

where P_N is the noise power measured over a bandwidth Δf . For a resistor, the noise temperature corresponds to its physical temperature. For other components, the noise temperature does not necessarily correspond to a physical temperature. Instead, it describes the physical temperature of a resistor that would produce an equivalent Johnson-Nyquist noise power.

Previously we restricted ourselves to the low-frequency limit ($hf \ll k_B T$). The more general expression for Johnson-Nyquist noise valid at all frequencies, expressed as a noise power, is given by

$$P_N = \int \frac{hf}{e^{hf/k_B T} - 1} df \quad (7.4)$$

where the integral is evaluated over the measurement bandwidth. Using a Taylor series for the exponential term, one can show that this reduces to $P_N = k_B T \Delta f$ when $hf \ll k_B T$. If we remember our statistical physics, we notice that the integrand is the Planck distribution describing the photon distribution function multiplied by the photon energy hf . This is because Johnson-Nyquist noise is just blackbody radiation from a one-dimensional source.

7.3 SHOT NOISE

Shot noise arises from the fact that charge is quantized. The DC current describes the average rate at which electrons pass through some point in a circuit. There is a corresponding average time between each electron passing through this point, but the actual time between two successive electrons can be smaller or larger than this average. The statistical variation of these times is described by a Poisson distribution, and it gives rise to Shot noise. The phenomenon was first observed in 1918 by Walter Schottky in studies of the current in vacuum tubes.

Unlike Johnson-Nyquist noise, Shot noise is only present if there is a nonzero average current. Shot noise is also diminished in systems in which electrons undergo multiple internal scattering events. This is because internal scattering averages out the statistics of the electron transit times. As a result, macroscopic resistors do not exhibit Shot noise. Besides vacuum tubes, another type of component that has Shot noise is one whose resistance arises from quantum tunneling through a potential barrier. Such a component can be made by placing a very thin ($\sim$ nm) insulating layer between two metallic layers; this is known as a tunnel junction. Another example is a reverse-biased PN junction, in which electrons can tunnel through the depletion region at the PN interface.

For a tunnel junction with a resistance R , the voltage noise spectral density due to Shot noise is given by

$$S_V = 2q_0 V R \quad (7.5)$$

where V is the DC voltage applied across the junction and q_0 is the fundamental charge. Since not all resistors exhibit the full Shot noise of Equation 7.5, we can describe the measured Shot noise of a resistor by the modified equation $S_V = 2F q_0 V R$ where F , known as the Fano factor, is a unitless quantity with a value between 0 and 1. The value of F contains information about how charge passes through the component.

Since a tunnel junction has resistance, it will also exhibit Johnson-Nyquist noise. There is a general expression describing the contributions of both Shot and

Johnson-Nyquist noise in such a system:

$$S_V = 2q_0VR \coth\left(\frac{q_0V}{2k_BT}\right) \quad (7.6)$$

Equation 7.6 reduces to Equation 7.5 (Shot noise) when $q_0V \gg k_BT$ and to Equation 7.2 (Johnson-Nyquist noise) when $q_0V \ll k_BT$.

7.4 AMPLIFIERS

In [section 7.2](#) we considered the effect of the impedance of the measurement circuit on a measurement of Johnson-Nyquist noise. Of course, the measurement circuit has its own noise. Here we will consider how to account for the noise added by an amplifier. All amplifiers add some noise to a measurement, and by convention we specify the noise referred to the amplifier's input.

First we will consider the case where $R_{in} \gg R$, where R_{in} is the input impedance of the amplifier and R is the impedance of the system connected to the amplifier input. This is the typical low-frequency case, where we want a very large input impedance to maximize voltage transfer. (Here we are assuming for simplicity purely real impedances, although one can generalize the approach to complex impedances.) If the amplifier produces some noise $\langle \delta V_{amp}^2 \rangle$ referred its input, then the resulting noise at its output will be $\langle \delta V_{out}^2 \rangle = |H|^2 \langle \delta V_{amp}^2 \rangle$, where $|H|$ is the magnitude of the amplifier's transfer function. If the system that is being measured has its own noise $\langle \delta V_{sys}^2 \rangle$, then this will add to $\langle \delta V_{amp}^2 \rangle$ and produce a total noise at the output of $\langle \delta V_{out}^2 \rangle = |H|^2 (\langle \delta V_{sys}^2 \rangle + \langle \delta V_{amp}^2 \rangle)$.

For high input-impedance amplifiers, it is common to specify the sensitivity in terms of the square root of the noise voltage spectral density $S_{V,amp}^{1/2}$ and the square root of the noise current spectral density $S_{I,amp}^{1/2}$. For a voltage amplifier, one must still consider the effect of current noise, as current noise is converted into voltage noise by the resistance R between the amplifier input and ground, i.e. $\delta V = R\delta I$. So the total voltage noise due to the amplifier at the amplifier input is $\langle \delta V_{amp}^2 \rangle = (S_{V,amp}^{1/2} + S_{I,amp}^{1/2}R)^2 \Delta f$.

Exercise 7.1: Measuring Johnson-Nyquist Noise with a High-Impedance Voltage Amplifier

A high-impedance voltage amplifier has $|H| = 500$, $S_{V,amp}^{1/2} = 15 \text{ nV/Hz}^{0.5}$, $S_{I,amp}^{1/2} = 0.18 \text{ pA/Hz}^{0.5}$, and a measurement bandwidth $\Delta f = 1.0 \text{ MHz}$. A $100 \text{ k}\Omega$ resistor at room temperature (296 K) is connected between the amplifier input and ground. What is $\langle \delta V_{out}^2 \rangle$ measured at the amplifier output?

Solution: The amplifier noise is $\langle \delta V_{amp}^2 \rangle = (S_{V,amp}^{1/2} + S_{I,amp}^{1/2}R)^2 \Delta f = 1.089 \times 10^{-9} \text{ V}^2$. This adds with the Johnson-Nyquist noise of the

resistor, which is $\langle \delta V_{JN}^2 \rangle = 4k_B T R \Delta f = 1.634 \times 10^{-9} \text{ V}^2$. We multiply the sum by $|H|^2$ to refer it to the output, giving $\langle \delta V_{out}^2 \rangle = 6.8 \times 10^{-4} \text{ V}^2$.

Next we consider the case where $R_{in} = R$, which is typical for microwave-frequency amplifiers. (Again we assume for simplicity purely real impedances.) We define P_{amp} as the amplifier noise power referred to the amplifier's input. It is common to specify the sensitivity of a microwave amplifier in terms of a noise temperature $T_{amp} = P_{amp} / (k_B \Delta f)$. It can also be expressed as a noise figure (NF) in decibels (dB), where $NF = 10 \log(T_{amp} / 290 + 1)$. If we are measuring an impedance-matched system with its own noise power P_{sys} , then the total noise power at the amplifier input is $P_{in} = P_{amp} + P_{sys}$. The total noise power at the amplifier output is then $P_{out} = G_P P_{in}$ where G_P is the amplifier power gain. If the system impedance R is not exactly matched to the amplifier input impedance R_{in} , then we can use the reflection coefficient $\Gamma = (R_{in} - R) / (R_{in} + R)$ from [chapter 4](#) to modify our expression for P_{in} as $P_{in} = P_{amp} + (1 - |\Gamma|^2) P_{sys}$.

Exercise 7.2: The Y-Factor Measurement

The Y-factor measurement is a technique for determining the gain and noise temperature of a microwave-frequency amplifier. It involves measuring the output noise power with an impedance-matched resistor at the amplifier input for two different physical temperatures of the resistor. For this exercise, consider a Y-factor measurement in which the output noise measured from 2.00 to 3.00 GHz is $5.84 \times 10^{-8} \text{ W}$ when the input resistor is at 296 K and $4.33 \times 10^{-8} \text{ W}$ when the input resistor is at 77.4 K. What are the amplifier noise temperature T_{amp} and power gain $G_P = P_{out} / P_{in}$?

Solution: We can use the equation for the total output power due to both amplifier noise and the Johnson-Nyquist noise of the resistor $P_{out} = G_P k_B (T_{amp} + T_R) \Delta f$ where T_R is the physical temperature of the resistor and $\Delta f = 1.00 \text{ GHz}$ is the measurement bandwidth. Solving this for G_P , we get $G_P = P_{out} / (k_B (T_{amp} + T_R) \Delta f)$. We can then write versions of this equation for the two different cases, equate these two equations, and then solve for T_{amp} . We get $T_{amp} = 550 \text{ K}$. We can then plug this value back in to one of our equations for G_P and we get $G_P = 5,000$.

Note that, because amplifiers add their own noise to a measurement, they *always* degrade the S/N of the measurement. This may seem counterintuitive, since amplifiers are commonly used in measurements where the signal amplitude is very small. In these situations, we use amplifiers to boost the signal amplitude to a level that is easily measured by an instrument such as a voltmeter, an oscilloscope, or a spectrum analyzer. These instruments have their own noise, and the job of the amplifier is to boost the signal amplitude so it is significantly larger than the instrument noise.

7.5 OTHER NOISE SOURCES

Many physical systems exhibit a noise spectrum that, at sufficiently low frequency, exhibits a spectral density that is proportional to $f^{-\alpha}$, where f is the frequency and α is a positive number approximately equal to one. This is often called $1/f$ (“one over f ”) noise or flicker noise. The frequency below which this behavior dominates is known as the $1/f$ knee. One generally wants to operate an AC circuit at frequencies above the $1/f$ knee in order to avoid excessive noise. $1/f$ noise also explains why, in general, AC measurements are almost always better than DC measurements when high sensitivity is desired.

The origins of $1/f$ noise depend on the physical system. Its ubiquity across a wide range of physical systems – from the resistance of resistors to the luminosity of stars to the water level at high tide – suggests that it comes from some fundamental truth about statistics. This was once described succinctly by a former professor of mine, whom I paraphrase: If you wait long enough, eventually something really bad will happen.

Another common source of noise is pickup noise. This arises when your measurement captures an unwanted signal from some other source. Perhaps the most common source of pickup noise is the electrical power line, which in the US has a frequency of 60 Hz. It is common to have a small amount of 60 Hz noise on the output of a DC power supply, which is known as *ripple*. When performing sensitive AC measurements, it is a good idea to stay away from 60 Hz. It is also a good idea to stay away from the first handful of its harmonics (integer multiples of 60 Hz), as nonlinearities can generate harmonics. Pickup noise can also come from other sources such as radio stations, wifi networks, and digital clock signals. Faraday shielding – i.e. putting your measurement apparatus inside a metal enclosure – can be an effective, if not always practical, way to minimize pickup noise from many types of sources. If you know the frequencies at which you are experiencing pickup noise, you can also try to filter out those frequencies.

7.6 SPECIAL TOPIC: DIGITAL FILTERING

The filters discussed in [section 3.2](#) are analog filters. A *digital* filter involves converting an analog signal into a digital signal and then applying a filter function to the signal using software rather than hardware. This can be done either in real time or after the data have been recorded. One advantage of digital filtering is that it allows filter functions that would be difficult or impossible to achieve with an analog filter. Digital filters can also be designed to adapt based on the signal parameters.

As a simple example, let’s imagine that we are measuring a detector that produces a voltage pulse with a fast rise and a slower exponential decay for each detection event. One measured pulse is shown in [Figure 7.3a](#). After looking at your data, you suspect there may be some pickup noise from the power line, so you take the Fourier transform of your pulse, which is shown to the right of the pulse. Your suspicion is confirmed - you see peaks characteristic of pickup noise from the power line at 60 Hz along with harmonics at 120 Hz and 180 Hz. To eliminate this pickup noise, you

apply a digital filter in which you just replace the amplitudes at each of these three frequencies with the average of the two adjacent data points, resulting in the Fourier transform seen in [Figure 7.3b](#). You then take the reverse Fourier transform, producing the digitally-filtered pulse shown to the left of the Fourier transform.

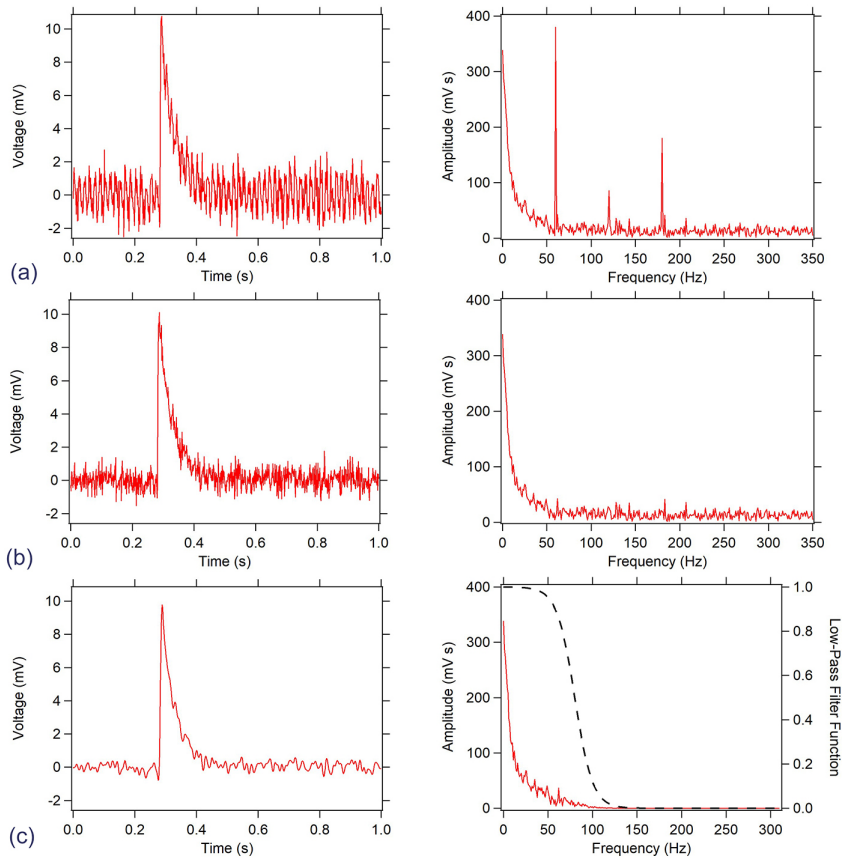


Figure 7.3 (a) Measured voltage pulse (left) and its Fourier transform (right). (b) Fourier transform after removing the pickup noise at 60 Hz, 120 Hz, and 180 Hz (right), along with the resulting pulse obtained from taking the reverse Fourier transform (left). (c) Fourier transform after applying the low-pass filter function shown in the black dashed line (right) and the resulting pulse obtained from taking the reverse Fourier transform (left).

This new pulse ([fig. 7.3b](#)) is clearly an improvement on the original pulse ([fig. 7.3a](#)). You might be happy with this improvement and call it a day. But, looking at the Fourier transformed data, you might notice that most of the signal from the pulse is below 50 Hz, while higher frequencies seem to contain mostly noise. Because of your newfound enthusiasm for digital filtering, you decide to apply a low-pass digital

filter function to reduce some of this high-frequency noise. (You want to be careful, however, since the fast pulse rise is contained in the higher frequency components of the signal, so if you get too aggressive with your low-pass filter you will start to reduce the pulse amplitude.) You multiply the Fourier-transformed data by the filter function $H(f) = 1/(1 + e^{(f-80)/10})$, where f is the frequency, which is shown as the black dashed line in Figure 7.3c. This figure also shows the Fourier-transformed data after the application of this filter function. Taking the reverse Fourier transform of this gives the pulse shown on the left side of fig. 7.3c. This is a much less noisy pulse than what you started with!

There are mathematical techniques for finding the optimal digital filter function for a particular application, but those details are beyond the scope of the present discussion. This simple illustration is meant to show that, while filtering with hardware is of course useful and important, there are instances in which one can get significant benefits from using digital filtering as part of the data analysis process.

7.7 PROBLEMS

Problem 1. This table contains the data from a time-domain oscilloscope waveform. Using these data, calculate the voltage noise spectral density S_V , which should have units of V^2/Hz . The measurement bandwidth extends from zero frequency to a maximum frequency given by $f_{max} = 1/(2\delta t)$, where δt is the time interval between successive data points. (You should calculate *one* value of S_V for the dataset.)

Time (ms)	Voltage (V)
0	2.27
0.1	2.21
0.2	2.19
0.3	2.30
0.4	2.25
0.5	2.32
0.6	2.30
0.7	2.20
0.8	2.22
0.9	2.27
1.0	2.21

Problem 2. A 50 kΩ resistor is connected between the input of a low-frequency voltage amplifier and ground. The amplifier is built from an OP-07 op-amp and has a transfer function $H = 1,000$. At the output of the amplifier, you measure $\langle \delta V^2 \rangle = 6.52 \times 10^{-5} V^2$ over a bandwidth $\delta f = 100$ kHz. If the op-amp’s noise voltage spectral density S_V and noise current spectral density S_I are constant over the measurement bandwidth, with $\sqrt{S_V} = 10$ nV/ $\sqrt{Hz}$ and $\sqrt{S_I} = 0.12$ pA/ $\sqrt{Hz}$, what is the physical temperature of the resistor at the amplifier input?

Problem 3. A $50\ \Omega$ microwave amplifier has a noise temperature $T_{amp} = 170\ \text{K}$ and a power gain of 900. If we connect a $50\ \Omega$ resistor at room temperature (296 K) to the amplifier input, what is the total noise power (in watts) at the amplifier output measured over the bandwidth from 4.0 to 8.0 GHz?

8 Digital Circuits

8.1 BINARY

Digital circuits represent information as binary numbers. Before discussing digital circuits, a brief review of binary numbers is in order. Binary is a base-2 number system in which each digit has two possible values, usually denoted as 0 and 1. This is in contrast to the familiar base-10 decimal number system, in which each digit has one of ten possible values (0 through 9). In both decimal and binary, a multi-digit number is written with the most-significant digit on the left and the least-significant digit on the right. As we all know, in decimal the least significant digit is the ones digit ($1 = 10^0$), the next least significant digit is the tens digit ($10 = 10^1$), the next is the hundreds digit ($100 = 10^2$), and so on. In this way, we know that the decimal number 472 means that we have 4 hundreds, 7 tens, and 2 ones. In binary, the least significant digit is the ones digit ($1 = 2^0$), the next least significant digit is the twos digit ($2 = 2^1$), the next is the fours digit ($4 = 2^2$), the next is the eights digit ($8 = 2^3$), and so on. Hence the binary number 1011 has 1 eight, 0 fours, 1 two, and 1 one, and its decimal equivalent is $8 + 2 + 1 = 11$.

A binary digit is known as a *bit*. An n -bit number has 2^n different possible values. Eight bits are equal to one *byte*. In binary, the prefixes kilo-, mega-, giga-, *etc.* differ not by the usual factor of 1,000 but rather by a factor of 1,024, because $1,024 = 2^{10}$. Hence 1,024 bytes are equal to 1 kilobyte (kB), and 1,024 kB are equal to 1 megabyte (MB). Note that B is the abbreviation for byte, while b is the abbreviation for bit.

In digital circuits, a binary zero is represented as a low voltage and a binary one is represented as a high voltage. Typically *low* is close to zero volts and *high* is close to five volts. The exact definition of what voltage corresponds to a binary zero and what voltage corresponds to a binary one depends on the digital logic family that one is using. There are a number of different digital logic families; two commonly-used ones are transistor-transistor logic (TTL) and 5V CMOS, where CMOS stands for *complementary metal-oxide-semiconductor*. The specified voltage ranges for a binary zero and a binary one for these two logic families are given in [Table 8.1](#). We note that the definitions are different for the input and the output of a circuit. Specifically, the output ranges are narrower than the input ranges. This is to build in protection from noise. If noise causes a change in the signal voltage between the output of one circuit and the input of another circuit, the larger range of the input will reduce the chances that the noise causes the voltage not to be read as the intended binary value.

Table 8.1
Comparison of TTL and 5V CMOS logic families

Binary Value	TTL		5V CMOS	
	Input	Output	Input	Output
0	0-0.8 V	0-0.5 V	0-1.5 V	0-0.05 V
1	2-5 V	2.7-5 V	3.5-5 V	4.95-5 V

8.2 ANALOG-TO-DIGITAL CONVERSION

Most signals in nature are analog, meaning that they are a continuously variable quantity. Most signal analysis is done on digital computers, which requires that we convert our measured analog signal into a digital signal for analysis. This is accomplished using an analog-to-digital (A/D) converter. To do this conversion, one must choose a sampling rate, which is the inverse of the time interval at which the analog signal is measured. For example, a sampling rate of 1 kHz corresponds to measuring the signal once every millisecond. We also have to choose how many bits we will use to represent each measured value. The size of the resulting digital file can be found by multiplying the sampling rate by the number of bits per sample by the total duration of the analog signal;¹ an example of this is given in [exercise 8.1](#).

Exercise 8.1: A/D Conversion

An analog audio-frequency signal is sampled at a rate of 20 kHz and each value is represented as an 8-bit number. If the total duration of the signal is 6 minutes, what is the resulting digital file size?

Solution:

$$\left(\frac{20 \times 10^3 \text{ samples}}{\text{s}}\right) \left(\frac{8 \text{ bits}}{\text{sample}}\right) (360 \text{ s}) = 5.76 \times 10^7 \text{ bits}$$

Dividing by 8 to convert from bits to bytes, we get 7.2×10^6 bytes. Then, if we divide by 2^{20} to convert from bytes to MB, we get 6.87 MB.

Because A/D conversion involves mapping a continuous variable onto a discrete variable, there is invariably a loss of information that results. This information loss can be minimized by using a larger number of bits per sample and by increasing the

¹Here we consider the raw output of A/D conversion. In practice, compression algorithms are often used to reduce the size of the resulting digital file.

sampling rate, but it cannot be totally eliminated. One key advantage to digitizing information is that it is possible to copy digital information with no loss of fidelity. Copies of analog information are always imperfect, and that imperfection grows if one makes a copy of a copy. (Think of making a photocopy of a document, then photocopying that photocopy, and then photocopying that photocopy in turn. Each time the process repeats, the document quality is reduced.) But digital information, due to its discrete nature, can be copied exactly an arbitrary number of times.

8.3 LOGIC GATES

The processing of binary information in a digital circuit is accomplished using digital logic gates. The behavior of a particular logic gate is described by a truth table, which lists the binary output for each binary input combination. Commonly used logic gates and their corresponding truth tables are given in [Table 8.2](#).

In the logic gate circuit symbols, only the input and output connections are shown; the power supply, usually denoted V_{cc} , is not shown. Since logic gates are active circuits, they will not function without an appropriate power supply. On an actual logic gate chip, the V_{cc} pin should be connected to an appropriate supply voltage – usually 5 V – and the ground pin should be connected to the circuit ground in order for the gate to function.

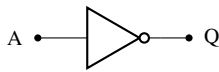
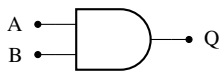
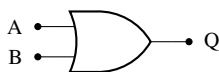
Logic gates are built from components such as transistors and resistors. An example of a simple AND gate circuit is shown in [Figure 8.1](#). (A commercial AND gate circuit will likely be more complex than this example, in order to achieve improved performance.) We can see that applying 5 V (a binary 1) to both inputs A and B will result in a base current flowing into each of the NPN bipolar junction transistors. This in turn increases the current that flows from V_{cc} , a fixed DC voltage, through each transistor and then through resistor R_2 to ground, producing a voltage at the output Q that is equal to this current multiplied by R_2 . Setting A or B to 0 V (a binary 0) will significantly decrease the current flowing from V_{cc} through the transistors to ground, resulting in a much smaller voltage at Q .

When building digital circuits using logic gates, it is not common to build the gates oneself from fundamental components. It is much more common to use commercial gate integrated circuits, which are relatively inexpensive, and just assume that the gate operates according to its truth table without worrying about the internal details.

We can combine logic gates to form more complex digital circuits known as combinational logic circuits. We determine the truth table for a combinational logic circuit from the truth tables of the logic gates that comprise the circuit. A simple example is shown in [Figure 8.2](#), in which two NAND gates are combined to make a single AND gate. If both A and B are 1, the first NAND gate outputs a 0 to both inputs of the second NAND gate, which in turn outputs a 1. Any other input combination will produce a 1 output from the first NAND gate and hence a 0 from the second NAND gate.

The operation of each logic gate has a corresponding function in the formalism of Boolean algebra. These individual gate functions can be combined to create Boolean

Table 8.2
Common digital logic gates

Name	Circuit Symbol	Truth Table															
NOT		<table><tr><th>A</th><th>Q</th></tr><tr><td>0</td><td>1</td></tr><tr><td>1</td><td>0</td></tr></table>	A	Q	0	1	1	0									
A	Q																
0	1																
1	0																
AND		<table><tr><th>A</th><th>B</th><th>Q</th></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	A	B	Q	0	0	0	0	1	0	1	0	0	1	1	1
A	B	Q															
0	0	0															
0	1	0															
1	0	0															
1	1	1															
NAND		<table><tr><th>A</th><th>B</th><th>Q</th></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>0</td></tr></table>	A	B	Q	0	0	1	0	1	1	1	0	1	1	1	0
A	B	Q															
0	0	1															
0	1	1															
1	0	1															
1	1	0															
OR		<table><tr><th>A</th><th>B</th><th>Q</th></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	A	B	Q	0	0	0	0	1	1	1	0	1	1	1	1
A	B	Q															
0	0	0															
0	1	1															
1	0	1															
1	1	1															
NOR		<table><tr><th>A</th><th>B</th><th>Q</th></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td></tr></table>	A	B	Q	0	0	1	0	1	0	1	0	0	1	1	0
A	B	Q															
0	0	1															
0	1	0															
1	0	0															
1	1	0															
XOR		<table><tr><th>A</th><th>B</th><th>Q</th></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>0</td></tr></table>	A	B	Q	0	0	0	0	1	1	1	0	1	1	1	0
A	B	Q															
0	0	0															
0	1	1															
1	0	1															
1	1	0															
XNOR		<table><tr><th>A</th><th>B</th><th>Q</th></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	A	B	Q	0	0	1	0	1	0	1	0	0	1	1	1
A	B	Q															
0	0	1															
0	1	0															
1	0	0															
1	1	1															

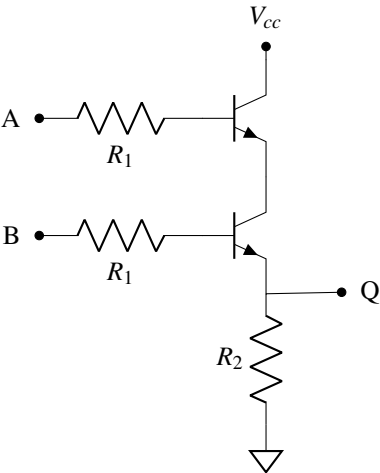


Figure 8.1 Simple realization of an AND gate made from two NPN bipolar junction transistors.

functions for combinational logic circuits composed of multiple gates. We will not cover Boolean algebra here, but the curious student can always look up more information on their own.

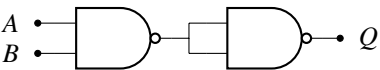


Figure 8.2 AND gate made from two NAND gates.

As another example, consider the circuit shown in [Figure 8.3](#), which is known as a full adder. Note that, due to the complexity of the circuit, we have adopted the convention here that two crossing wires are only connected if a dot is drawn at the point of crossing; if there is no dot, the crossing wires are not connected. The inputs to the circuit are three one-bit numbers A , B , and C_i , and the output is a two-bit number C_oS , where C_oS is equal to the sum of the three inputs. C_i is called the input carry bit, C_o is called the output carry bit, and S is called the sum bit. The truth table for this circuit is given in [Table 8.3](#); it can be verified that this truth table is consistent with the truth tables of the individual logic gates.

We can go up another level in the digital circuit hierarchy and represent our full adder circuit as a single component, as seen in [Figure 8.4](#). We can then string together N full adder circuits to create a circuit that adds two N -bit binary numbers. An example with $N = 4$ is shown in [Figure 8.5](#), where the inputs are the two four-bit numbers $A_4A_3A_2A_1$ and $B_4B_3B_2B_1$ and the output, their sum, is the five-bit number

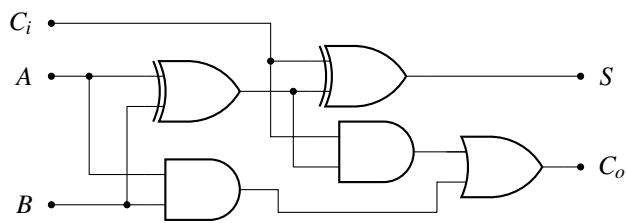


Figure 8.3 Full adder circuit. Note that crossing wires are only connected if a dot is drawn at the point of crossing.

Table 8.3
Truth table for the full adder circuit

<i>A</i>	<i>B</i>	<i>C_i</i>	<i>C_o</i>	<i>S</i>
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
1	0	0	0	1
0	1	1	1	0
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

$O_5O_4O_3O_2O_1$. Note that the output carry bit from one stage passes to the input carry bit in the next stage.

8.4 DIGITAL MEMORY CIRCUITS

Digital memory circuits are composed of multiple logic gates with some connection between the circuit output and input, i.e. with feedback. An example is the set-reset (SR) latch circuit, sometimes called a SR flip-flop, which can be built using two NOR gates, as shown in Figure 8.6. *S* is the set input bit, *R* is the reset input bit, and *Q* and $\overline{Q}$ are the output bits. In Boolean algebra, a bar over a variable indicates that it is the opposite of the variable without the bar, hence we know from the labels that the outputs *Q* and $\overline{Q}$ should always be opposite, i.e. if *Q* is a binary 0 then $\overline{Q}$ should be a binary 1 and vice versa.

Because of the feedback, it is not possible to specify a particular input and then determine the resulting output using the individual gate truth tables. In analyzing digital memory circuits, it is common to do a sequence analysis. In this process, one assumes particular values of both the inputs and outputs, verifies that these are consistent with the individual gate truth tables, and then steps through a sequence

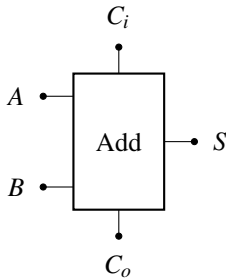


Figure 8.4 Component representation of a full adder.

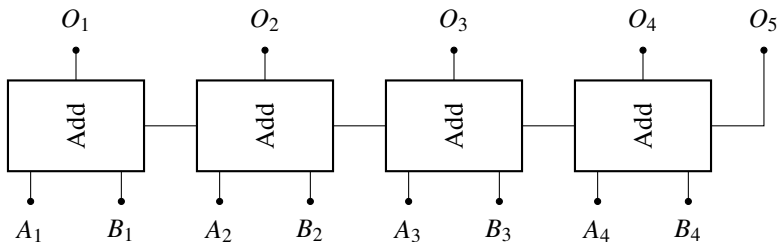


Figure 8.5 Circuit that adds two 4-bit numbers $A_4A_3A_2A_1$ and $B_4B_3B_2B_1$ to produce their 5-bit sum $O_5O_4O_3O_2O_1$.

of input changes to see the resulting outputs. In this way, one can determine the truth table for the circuit. Note that it is possible to have more than one valid output state for a particular input, and it is also possible to have no valid output state for a particular input – this is sometimes called a metastable state.

It is also possible to do a brute force analysis in which, for each input combination, each possible output combination is checked for validity based on the NOR truth table. For example, for the input combination $S,R = 0,0$, there are four possible output combinations: $Q, \overline{Q} = 0,0; 0,1; 1,0$; and $1,1$. Given the notation of the output variables as Q and $\overline{Q}$, we expect that $0,0$ and $1,1$ are not valid outputs, but we can check to be sure. When we check all four possible output combinations for each of the four input combinations, we find the following: When $S,R = 0,0$, both $Q, \overline{Q} = 0,1$ and $1,0$ are valid. When $S,R = 0,1$, only $Q, \overline{Q} = 0,1$ is valid. When $S,R = 1,0$, only $Q, \overline{Q} = 1,0$ is valid. And when $S,R = 1,1$, there are no valid outputs.

To understand what is happening for $S,R = 0,0$, we can do a sequence analysis. We'll begin with $S,R = 0,1$ and $Q, \overline{Q} = 0,1$, which we previously found was a valid state. Then we switch R from 1 to 0 and propagate this through the circuit. To this, we first update the output of the top NOR gate, then feed back any change in its output Q to the bottom NOR gate, and finally we update the output of the bottom

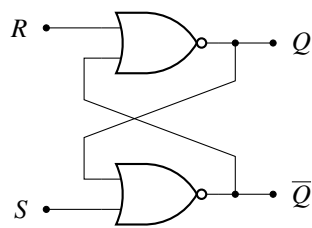


Figure 8.6 Set-reset (SR) latch circuit.

Table 8.4
Truth table for the SR latch circuit

S	R	Q	$\overline{Q}$
0	0	Q_{n-1}	$\overline{Q}_{n-1}$
0	1	0	1
1	0	1	0

NOR gate as needed. What we find is that changing R from 1 to 0 does not change the output Q of the top NOR gate, and hence $\overline{Q}$ also remains unchanged. Next we can consider starting from the state $S, R = 1, 0$ and $Q, \overline{Q} = 1, 0$, which is also valid, and then switching the S from 1 to 0. This produces no change in the output $\overline{Q}$ of the bottom NOR gate, and hence Q will also remain the same. What we find is that however we get to $S, R = 0, 0$, the outputs will remain the same as the previous step. Thus we call $S, R = 0, 0$ a *hold* state; it is also known as a *latched* state. We can then write the truth table for the SR latch as shown in Table 8.4.

What we see is that a 1 at the set bit will set the value of Q to 1, while a 1 at the reset bit will reset Q to zero. The notation Q_{n-1} and $\overline{Q}_{n-1}$ means that the values of Q and $\overline{Q}$ at step n will remain the same as the previous step; it is one way of designating a hold state. The metastable state of $S, R = 1, 1$, for which there are no valid outputs, is usually left off the truth table. In designing a circuit with a SR latch, one should avoid this particular input combination. Sometimes the $\overline{Q}$ state is left off the truth table, since we know it should always be the opposite of Q .

We can go up a level in the circuit hierarchy and represent the SR latch as a discrete component, as shown in Figure 8.7.

One application of the SR latch is debouncing a mechanical switch. Switch bounce is a phenomenon that commonly occurs when closing a mechanical switch. When the switch closes, one piece of metal is rapidly brought into contact with another piece of metal. The collision between these two pieces of metal is not perfectly inelastic, and hence they tend to bounce off each other several times before continuous contact is made.

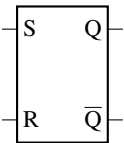


Figure 8.7 Component representation of SR latch.

As an example, consider the circuit shown in [Figure 8.8](#). Here a 5.0 V DC source is connected to a 1 kΩ resistor through a mechanical switch. The time-dependent voltage across the resistor immediately following the closing of the switch at time $t = 0$ is shown in [Figure 8.8](#), in which the effect of switch bounce is clearly seen.

We can debounce the switch by connecting an SR latch between the switch and the resistor, with the S input connected to the switch and the Q output connected to the resistor, as shown in [Figure 8.9](#). We see that this eliminates the switch bounce seen in previous figure. One limitation is that this requires a signal that is consistent with a binary 1; it will not work for voltages outside this range.

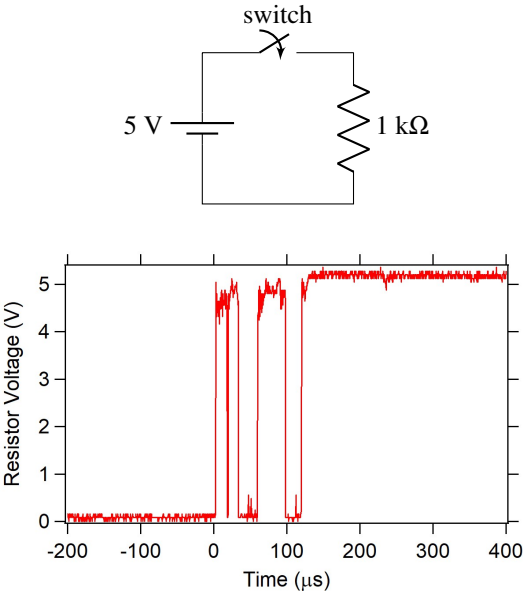


Figure 8.8 Example of a circuit showing switch bounce.

We can replace the NOR gates of our SR latch with NAND gates to create a $\overline{S}\overline{R}$ latch, whose schematic and truth table are shown in [Figure 8.10](#). This truth table is the same as the SR latch but with the values of the S and R bits inverted.

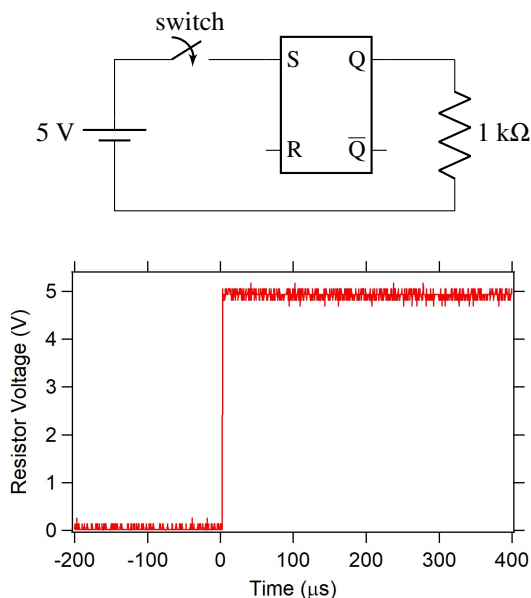


Figure 8.9 Debounced switch using SR latch.

8.5 CLOCKED DIGITAL CIRCUITS

A digital clock is a square wave that alternates between 0 V (a binary 0) and 5 V (a binary 1) with a 50% duty cycle. (The duty cycle of a square wave is the percentage of time spent in the high state.) In clocked digital circuits, the clock frequency sets the rate at which information is processed by the circuit. Most clocked digital circuits are what is known as *edge-triggered*, meaning that the inputs are read in and the outputs updated accordingly on each clock edge, where the clock edge refers to when the clock changes.

We can make a clocked version of our SR latch circuit as shown in [Figure 8.11](#). In this circuit, if the clock (*CLK*) signal is 0, then the outputs of the two AND gates will be 0 regardless of the values of *S* and *R*, and hence the outputs will hold. When the clock is 1, the *R* and *S* inputs will be fed into each NOR gate, with the outputs responding accordingly. This is sometimes called a *gated* SR latch, as it behaves as a regular SR latch when the clock is high (1) and does nothing when the clock is low (0).

Another example of a clocked digital circuit is the delay (D) flip-flop, whose circuit schematic and component representation are shown in [Figure 8.12](#). In the component representation, the input marked with the triangle is the clock input. More specifically, the triangle indicates that this is a *rising-edge* triggered circuit, i.e. the outputs only change when the clock rises from 0 to 1. (A falling-edge triggered circuit is often indicated by placing a small circle, denoting an inversion, immediately

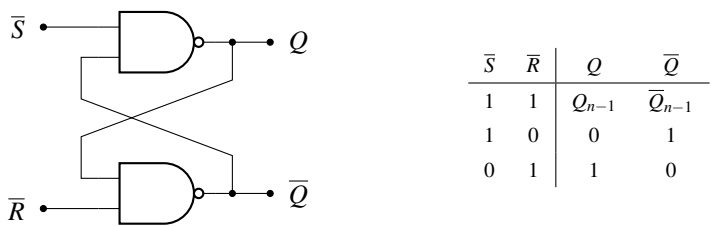


Figure 8.10 $\bar{S}\bar{R}$ latch circuit and its truth table.

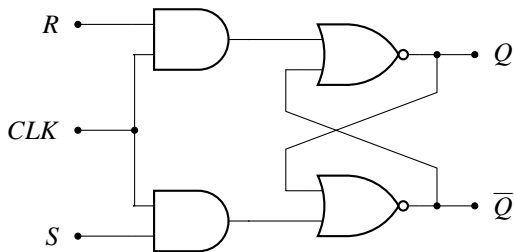


Figure 8.11 Clocked SR latch circuit.

in front of the triangle clock input.) In the schematic, the two right-most NAND gates function as an $\bar{S}\bar{R}$ latch whose truth table is given in [Figure 8.10](#). If the clock is 0, then the first two NAND gates will each output 1 regardless of D , resulting in a hold state at the outputs Q and $\bar{Q}$. If the clock is 1 and D is 0, then the top left NAND outputs a 1 and the bottom left NAND outputs a 0, resulting in $Q = 0$. If the clock is 1 and D is 1, then the top left NAND outputs a 0 and the bottom left NAND outputs a 1, resulting in $Q = 1$. This results in the truth table given in [Table 8.5](#). In this truth table, the designation X means that the value doesn't matter, i.e. it can be either a 0 or a 1. We can summarize the truth table by saying that the outputs hold when the clock is low (0) and D is written to Q when the clock is high (1).

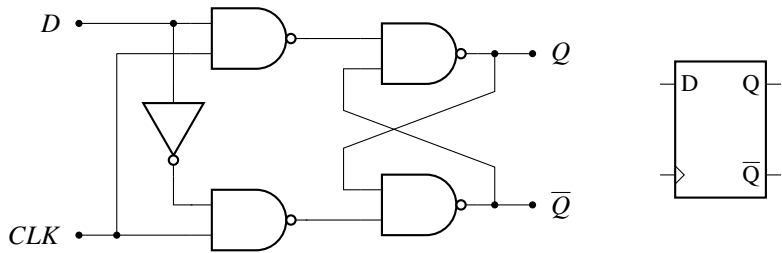


Figure 8.12 Circuit schematic (left) and component representation (right) of D flip-flop.

Table 8.5
Truth table for the D flip-flop

<i>CLK</i>	<i>D</i>	<i>Q</i>	$\overline{Q}$
0	X	Q_{n-1}	$\overline{Q}_{n-1}$
1	0	0	1
1	1	1	0

It is possible to combine multiple D flip-flops to make a circuit called a shift register. A shift register is useful for storing and moving binary information in a time-controlled fashion and is also known as a delay line memory. A four-bit shift register made from four D flip-flops is shown in Figure 8.13. This circuit takes a data input bit and, on the next rising clock edge, writes this value to the first bit in the shift register Q_1 . On the next rising clock edge, this Q_1 value is written to the next bit in the register Q_2 and Q_1 is updated with the current value of the data input. On each rising clock edge, the data in the register moves one step to the right, i.e. Q_{n+1} takes on the previous value of Q_n . The data in the right-most position in the register – Q_4 in this example – is discarded on the next clock rising edge. One can expand the shift register by simply adding more D flip-flops.

Another example of a clocked digital memory circuit is the JK flip-flop, shown in Figure 8.14. In this schematic, we have introduced a new component: the three-input NAND gate. This gate outputs a 1 unless all three of its inputs are 1, in which case it outputs a zero. (An exercise for the motivated reader: How can you build a three-input NAND gate out of standard two-input NAND gates?)

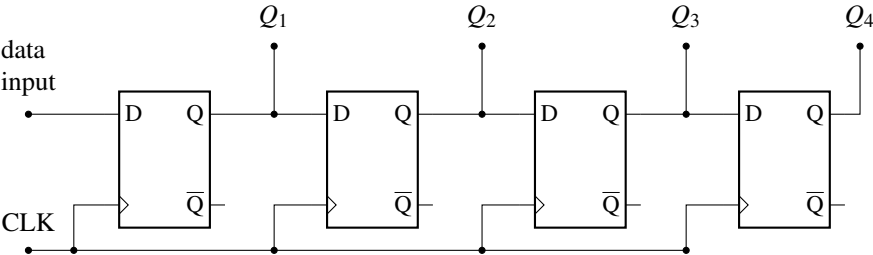


Figure 8.13 Four-bit shift register.

As in our D flip-flop, the two right-most NAND gates in our JK flip-flop comprise an $\overline{SR}$ latch. When the clock is 1, both three input NANDs are presented with a 0 at at least one of their inputs, resulting in a 1 at their output. This produces a hold state at the output of our $\overline{SR}$ flip-flop, regardless of the values of J and K . When

Table 8.6
Truth table for the falling-edge JK flip-flop

CLK	J	K	Q	$\overline{Q}$
1	X	X	Q_{n-1}	$\overline{Q}_{n-1}$
0	0	0	Q_{n-1}	$\overline{Q}_{n-1}$
0	0	1	0	1
0	1	0	1	0
0	1	1	$\overline{Q}_{n-1}$	Q_{n-1}

the clock is 0, the outputs do depend on J and K , as summarized in Table 8.6. The output $Q, \overline{Q} = \overline{Q}_{n-1}, Q_{n-1}$ is known as a toggle state, which means that Q will be whatever $\overline{Q}$ was in the previous step and vice versa. More specifically, this is known as a *falling-edge* JK flip-flop, because the output always holds on the rising clock edge. (We can convert this to a *rising-edge* JK flip-flop simply by removing the NOT gate in Figure 8.14.)

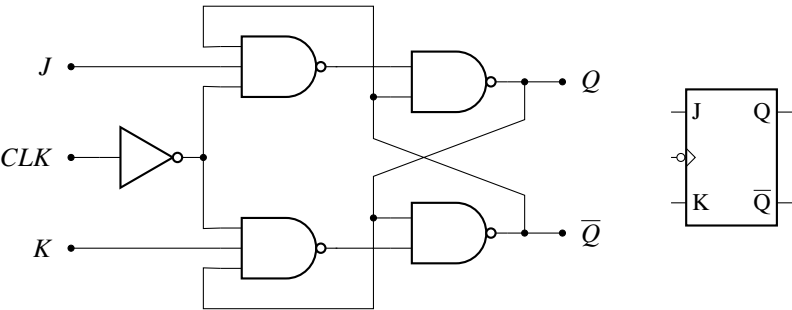


Figure 8.14 Circuit schematic (left) and component representation (right) of a falling-edge JK flip-flop.

We can combine multiple JK flip-flops to make a binary counter circuit, a four-bit version of which is shown in Figure 8.15. This circuit produces a four-bit output $Q_4Q_3Q_2Q_1$ that begins at 0000 and, on the first falling clock edge, increases by one to 0001. On the next falling clock edge it increases to 0010, then to 0011, then to 0100, and so on all the way to 1111. On the next falling clock edge after this, it starts over at 0000. In other words, it counts in binary from a decimal equivalent of 0 to a decimal equivalent of 15 on an infinitely repeating loop.

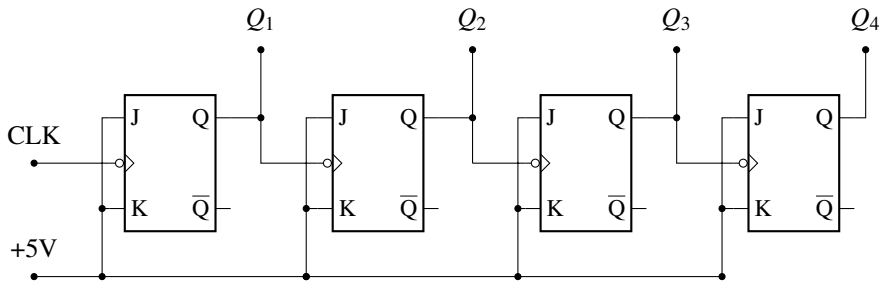


Figure 8.15 Four-bit binary counter.

8.6 BUILDING BLOCKS OF A COMPUTER

We have gone several steps up the digital circuit hierarchy, starting from logic gates, then using these to build combinational logic circuits and flip-flop circuits, and then using flip-flop circuits to build more complex circuits such as the shift register and the binary counter. Continuing up the digital circuit hierarchy all the way to a complete digital computer is beyond the scope of this text, but we can briefly outline the key ingredients.

The “brain” of a computer is the processor, also known as a central processing unit or CPU. An essential component of a CPU is the arithmetic logic unit (ALU). This is a circuit built from logic gates that can perform binary arithmetic – including addition and subtraction – as well as bitwise logical operations (the functions of individual logic gates). Besides the ALU, a CPU also requires a control unit (CU) and memory. The CU directs the operation of the CPU, providing instructions to the ALU and coordinating the inputs and outputs. The memory, as you can probably guess, stores digital information until it is needed.

The first commercial microprocessor – a CPU microfabricated on a single chip – was the Intel 4004, released in 1971. It contained a total of 2,250 transistors, had sixteen 4-bit registers, and ran on a 740 kHz clock frequency. Intel released its first Pentium CPU in 1993 with 3.1 million transistors and a 60 MHz clock frequency. Today’s CPUs have in excess of a billion transistors and clock frequencies of several GHz. For further details on CPU design, the interested reader should consult a book on computer architecture.

8.7 SPECIAL TOPIC: INTRODUCTION TO QUANTUM COMPUTING

8.7.1 CLASSICAL VERSUS QUANTUM BITS

Classically, a bit contains one piece of information: whether it is a 0 or a 1. A quantum bit, or qubit, is different than a classical bit in that it can exist not only in state 0 or state 1, it can also exist in a superposition of these two states. When measured, a qubit in such a superposition will have some probability of being found in state 0 and

some probability of being found in state 1. We can't directly measure the superposition; when we measure the system, we always measure it as being in either state 0 or state 1. But by repeating the measurement many times on an identically-prepared qubit, we can determine the properties of the superposition.

We can write an equation describing the state of our qubit ψ as a linear combination of the 0 state, given by ψ_0 , and the 1 state, given by ψ_1 :

$$\psi = A\psi_0 + B\psi_1$$

In quantum mechanics, ψ is known as the qubit's wavefunction. The probability of measuring the system in the 0 state is $|A|^2$ and the probability of measuring the system in the 1 state is $|B|^2$. Because the qubit must be in one of these two states, we must have $|A|^2 + |B|^2 = 1$. Here $|A|^2$ denotes the product of A and its complex conjugate, which ensures that, even if A is a complex number, the probability $|A|^2$ is always a real number.

We can write a wavefunction for two qubits in the form

$$\psi = A\psi_{00} + B\psi_{01} + C\psi_{10} + D\psi_{11}$$

where ψ_{00} refers to the state in which the first qubit is a 0 and the second qubit is a 0, which has a probability $|A|^2$; ψ_{01} refers to the state in which the first qubit is a 0 and the second qubit is a 1, which has a probability $|B|^2$; and so on. Here we have the requirement that $|A|^2 + |B|^2 + |C|^2 + |D|^2 = 1$. We see that this wavefunction is described by $2^2 = 4$ pieces of information, the values of the four coefficients A , B , C , and D . If we extend this to three qubits, then our wavefunction becomes

$$\psi = A\psi_{000} + B\psi_{001} + C\psi_{010} + D\psi_{011} + E\psi_{100} + F\psi_{101} + G\psi_{110} + H\psi_{111}$$

which has $2^3 = 8$ coefficients and hence requires 8 pieces of information to describe the state. As before, the sum of the complex conjugate squared of each coefficient must equal one.

Extrapolating from this, we see that, while a system of n classical bits contains n bits of information, a system of n qubits contains 2^n bits of information. As n gets large, the difference between the two cases becomes dramatic. For example, if we have 400 qubits, then our system contains $2^{400} \sim 10^{120}$ bits of information, which is significantly more than the estimated number of particles in the observable universe!

This exponential scaling of the information content with the number of qubits is what makes quantum computing so exciting. There are, however, some important caveats. Realizing the 2^n informational content of n qubits requires being able to produce entangled states, which we will discuss in the next section; without these entangled states, the information content of a system of n qubits only grows linearly with n , not exponentially. While making a system containing 400 qubits is relatively straightforward, generating an entangled state of 400 qubits is not so simple. At the

time of this writing, the largest system of entangled qubits that has been realized is around 20.²

Another important caveat is that these entangled states have a limited lifetime, a phenomenon known as decoherence. Additionally, gate operations (discussed in [section 8.7.3](#)) have a limited fidelity, i.e., they do not result in the desired outcome 100% of the time. The combination of finite qubit lifetimes and limited gate fidelities means that the result of a calculation becomes increasingly likely to be incorrect as the number of gate operations increases. One solution to this problem is known as quantum error correction. Quantum error correction uses an algorithm to detect and correct errors in real time. Of course, for such an algorithm to be beneficial, the rate at which it identifies and corrects errors must be greater than the rate at which it creates errors. This has not yet been achieved, but there is hope that, with continued improvements in qubit lifetimes and gate fidelities, it will be realized in the coming years.

Yet another caveat is that the informational content of a system of n qubits collapses to the classical value of n bits upon readout. This means that calculations on a quantum computer must be designed to make use of the exponentially larger informational capacity that comes from entangled states and then produce a final state that does not require this larger capacity prior to readout. Conventional algorithms are not designed to do this, meaning that there is no reason to expect a conventional algorithm to run faster on a quantum computer than on a classical computer. To take advantage of the abilities of a quantum computer, an entirely new class of quantum algorithms must be developed. For certain problems, such algorithms have already been developed – such as Shor’s algorithm for factoring prime numbers and Grover’s algorithm for unstructured search – and the development of new quantum algorithms is a growing field of research. But there is no indication that quantum algorithms will be useful for the majority of computational tasks. This means that quantum computers are not likely to replace classical computers; rather, they will likely supplement them for certain computational tasks.

8.7.2 HOW TO BUILD A QUBIT

To make a qubit, we need a quantum system that can be reliably placed in its lowest energy state, also known as its ground state, which represents a binary 0. We also need an energy difference between the ground state and the first excited state, which represents a binary 1, that is unique, such that when we give the system this energy we can be sure that it promotes the system from the ground state (0) to the first excited state (1) and not to some other state.

There are many quantum systems that can satisfy these requirements, and hence there are many possible types of qubits. These systems include trapped atoms, superconducting circuits, and spin-active defects in solid state systems. Most qubits are

²To help protect against errors, it is common to use multiple physical qubits to represent a single logical qubit. In general, the number of physical qubits is not the best metric by which to compare the capabilities of different quantum computers. One better metric is called the *quantum volume*, the details of which can be looked up by the curious reader.

either electrical, meaning that the 0-1 energy transition corresponds to photon frequencies in the microwave range, or optical, meaning that the 0-1 energy transition corresponds to photon frequencies in or near the visible range. To be able to reliably place the qubit in its ground state, the temperature of the qubit T must be such that $k_B T$ is much less than the energy difference between the 0 and 1 states, where k_B is Boltzmann's constant. For electrical qubits, this generally requires operation at temperatures below 0.1 K. For optical qubits, the cooling requirements are not so stringent. One advantage of electrical qubits is scalability, specifically the idea that one can apply existing tools and techniques for making large-scale microwave-frequency integrated circuits to making large-scale quantum integrated circuits.

One model for a qubit is the quantum anharmonic oscillator. A *harmonic* oscillator is a system with a quadratic potential. Examples include the mass on a spring, the pendulum (for small angle displacement), and, in electronics, the parallel LC circuit. The harmonic oscillator is a widely used model for low-energy systems in physics because the Taylor series expansion of some arbitrary potential for small deviations from some local minimum is approximately quadratic. When perturbed, a harmonic oscillator will oscillate at some characteristic frequency. For the mass on a spring, the characteristic frequency is $\omega = \sqrt{k/m}$, where k is the spring constant and m is the mass. For the LC circuit, we saw in [chapter 2](#) that the characteristic frequency is $\omega = 1/\sqrt{LC}$, also known as the resonant frequency.

In the quantum regime, our harmonic oscillator has some nonzero ground state energy. This arises because of the Heisenberg uncertainty principle. The energy difference between successive energy levels of the harmonic oscillator is the same and is equal to $\hbar\omega$, where $\hbar$ is the reduced Planck constant and ω is the characteristic angular frequency. The ground state is $\hbar\omega/2$. (A derivation of this can be found in most introductory quantum mechanics textbooks.)

An oscillator becomes *anharmonic* if the potential is not exactly quadratic. As a result, the energy difference between the successive energy levels is no longer the same. This is critical for a qubit, as it allows the 0 to 1 transition to be uniquely addressable. In an LC circuit, we can create this anharmonicity by using a nonlinear inductor or capacitor. One common approach for electrical qubits is to use a superconducting device called a Josephson junction, which acts as a nonlinear inductor. Placing this in parallel with a linear capacitor yields an electrical anharmonic oscillator and hence a qubit. One could also imagine making an anharmonic oscillator using a type of nonlinear capacitor called a varactor, which is a reverse-biased PN junction; this device makes use of the dependence of the width of the depletion region, and hence the junction capacitance, on the reverse-bias voltage. Such a nonlinear capacitor, placed in parallel with a linear inductor, also produces an anharmonic oscillator. The resistive losses, however, would yield a qubit with an extremely short lifetime. Superconducting circuits have very low losses and hence can produce relatively long qubit lifetimes.

8.7.3 QUANTUM GATES

The state of a qubit can be represented as a vector extending from the origin to a point on the surface of a three-dimensional unit sphere known as the Bloch sphere, shown in [Figure 8.16](#). On this sphere, the positive z -axis corresponds to the 0 state

and the negative z -axis corresponds to the 1 state. The orientation of the vector on the Bloch sphere can be specified by two pieces of information, the values of the angles θ and ϕ , with $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. Based on this picture, we can write the wavefunction for our qubit as

$$\psi = \cos\left(\frac{\theta}{2}\right)\psi_0 + e^{i\phi}\sin\left(\frac{\theta}{2}\right)\psi_1 \quad (8.1)$$

The probability of measuring our qubit in state 0 is $\cos^2(\theta/2)$ and the probability of measuring it in state 1 is $|e^{i\phi}\sin(\theta/2)|^2 = \sin^2(\theta/2)$.

A single-qubit gate is an operation that moves the state of our qubit from some initial location to some final location on the Bloch sphere. Three fundamental single-qubit gates are the Pauli-X gate, the Pauli-Y gate, and the Pauli-Z gate. The Pauli-X gate rotates the state of the qubit π radians about the x -axis and is analogous to a classical NOT gate. The Pauli-Y gate rotates the state of the qubit π radians about the y -axis. And the Pauli-Z gate rotates the state of the qubit π radians about the z -axis; it is also known as a phase-flip gate. Another important single-qubit gate is the Hadamard (H) gate, which produces a π rotation about the z -axis followed by a $\pi/2$ rotation about the y -axis (or, equivalently, a π rotation about the $(\hat{x} + \hat{z})/\sqrt{2}$ axis). The Hadamard gate takes a qubit that is initially in the 0 or the 1 state and puts it into a superposition state with equal probabilities of being measured as a 0 or a 1. In [Table 8.7](#), we summarize the result of applying each of these gates to the initial qubit state is given in Equation 8.1.

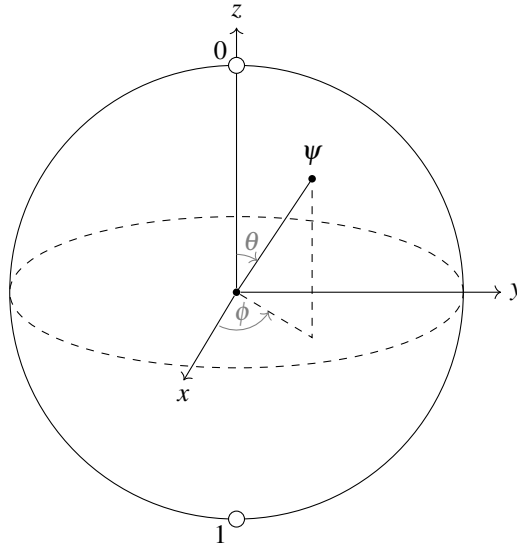


Figure 8.16 Bloch sphere representation of the single qubit state ψ .

Table 8.7**Results of applying single-qubit gates to the qubit state of Equation 8.1**

Gate	Resulting State
Pauli-X	$\psi = e^{i\phi} \sin\left(\frac{\theta}{2}\right) \psi_0 + \cos\left(\frac{\theta}{2}\right) \psi_1$
Pauli-Y	$\psi = e^{i\phi} \sin\left(\frac{\theta}{2}\right) \psi_0 - \cos\left(\frac{\theta}{2}\right) \psi_1$
Pauli-Z	$\psi = \cos\left(\frac{\theta}{2}\right) \psi_0 - e^{i\phi} \sin\left(\frac{\theta}{2}\right) \psi_1$
Hadamard	$\psi = \frac{1}{\sqrt{2}} \left[\left(\cos\left(\frac{\theta}{2}\right) + e^{i\phi} \sin\left(\frac{\theta}{2}\right) \right) \psi_0 + \left(\cos\left(\frac{\theta}{2}\right) - e^{i\phi} \sin\left(\frac{\theta}{2}\right) \right) \psi_1 \right]$

Now consider two qubits x and y , with the state of qubit x given by $\psi_x = \cos\left(\frac{\theta_x}{2}\right) \psi_0 + e^{i\phi_x} \sin\left(\frac{\theta_x}{2}\right) \psi_1$ and the state of qubit y given by $\psi_y = \cos\left(\frac{\theta_y}{2}\right) \psi_0 + e^{i\phi_y} \sin\left(\frac{\theta_y}{2}\right) \psi_1$. We can write the product $\psi_x \psi_y = \psi_{xy}$ as

$$\begin{aligned} \psi_{xy} = & \cos \frac{\theta_x}{2} \cos \frac{\theta_y}{2} \psi_{00} + \cos \frac{\theta_x}{2} e^{i\phi_y} \sin \frac{\theta_y}{2} \psi_{01} \\ & + e^{i\phi_x} \sin \frac{\theta_x}{2} \cos \frac{\theta_y}{2} \psi_{10} + e^{i(\phi_x + \phi_y)} \sin \frac{\theta_x}{2} \sin \frac{\theta_y}{2} \psi_{11} \end{aligned} \quad (8.2)$$

This state is specified by four parameters, θ_x , ϕ_x , θ_y , and ϕ_y . Likewise, if we consider three qubits x , y , and z , the product $\psi_x \psi_y \psi_z = \psi_{xyz}$ is described by six parameters, θ_x , ϕ_x , θ_y , ϕ_y , θ_z , and ϕ_z . But we said before that a three-qubit state contains $2^3 = 8$ bits of information, not $2 + 2 + 2 = 6$, i.e. the information content should grow exponentially with the number of qubits, not linearly. This is because a simple product of multiple individual qubit states does include entangled states. If the state of each qubit is specified independently and does not depend on the other qubits, then it is not an entangled state. In an entangled state, there must be some dependence of the state of one qubit on the other qubit(s).

In the previous section, we wrote a general expression for a two-qubit state as $\psi_{xy} = A\psi_{00} + B\psi_{01} + C\psi_{10} + D\psi_{11}$. If the four coefficients A , B , C , and D can be mapped onto values of θ_x , ϕ_x , θ_y , and ϕ_y in Equation 8.2, then it is not an entangled state. But if this is not possible, then it is an entangled state. Another way of saying this is that if $AD = BC$, then it is not an entangled state, but if $AD \neq BC$, then it is an entangled state. As the number of qubits n becomes large, the number of total possible states, which grows exponentially with n , becomes much larger than the number of nonentangled states, which grows linearly with n , and hence the vast majority of possible states for $n \gg 1$ are entangled states. This is why entanglement is key to realizing the potential of quantum computing.

As an example, the state $\psi_{xy} = (1/2)(\psi_{00} - \psi_{01} + \psi_{10} - \psi_{11})$ corresponds to Equation 8.2 with $\theta_x = \pi/2$, $\phi_x = 0$, $\theta_y = \pi/2$ and $\phi_y = \pi$, and hence this is not an entangled state. We can also verify that this state satisfies $AD = BC$. But if we

consider the state $\psi_{xy} = (\psi_{00} + \psi_{11})/\sqrt{2}$, we see that this cannot be described by Equation 8.2, and also $AD \neq BC$, and hence this is an entangled state. The physical meaning of this entangled state is that, if we measure just one of the two qubits, we have a 50% chance of finding it in state 0 and a 50% chance of finding it in state 1. But once we measure that qubit, the probability of a subsequent measurement of the other qubit yielding the same state becomes 100%.

Two-qubit gates act not on a single qubit but rather on a system of two qubits. One such two-qubit gate is the controlled-NOT (C-NOT) gate, which applies a Pauli-X gate to the second qubit if the first qubit is in the 1 state and does nothing if the first qubit is in the 0 state. If we consider the two-qubit state $\psi_{xy} = (1/2)(\psi_{00} - \psi_{01} + \psi_{10} - \psi_{11})$, applying a C-NOT gate to this state results in the new state $\psi_{xy} = (1/2)(\psi_{00} - \psi_{01} + \psi_{11} - \psi_{10})$. We see that the terms with the first qubit in state 0 are unchanged, while the terms with the first qubit in state 1 have the state of the second qubit flipped. One important quality of the C-NOT gate is its ability to convert between a nonentangled state and an entangled state. For example, applying the C-NOT gate to the nonentangled state $\psi_{xy} = (\psi_{00} + \psi_{10})/\sqrt{2}$, we get the entangled state $\psi_{xy} = (\psi_{00} + \psi_{11})/\sqrt{2}$. Applying the C-NOT gate again returns us to the original nonentangled state.

The Toffoli gate, also known as a controlled-controlled-NOT gate, is a three-qubit gate that applies a Pauli-X gate to the third qubit if and only if the other two qubits are both 1; otherwise, it does nothing. We can use a combination of Toffoli gates and C-NOT gates to build a quantum version of a full adder. First we need to introduce the circuit symbols for our gates. In quantum circuits, one represents each qubit as a horizontal line, with gates indicated along this line in chronological sequence from left to right. The C-NOT and Toffoli gates are shown in Figure 8.17, with each horizontal line representing one qubit. The quantum full adder is shown in Figure 8.18. This circuit requires five qubits. Three represent one bit numbers to be added, denoted A , B , and C_i . The other two are initialized in the zero state, and at the output these become the sum bit S and the output carry bit C_o representing the two-bit sum C_oS .



Figure 8.17 Circuit representation of C-NOT gate (left) and a Toffoli gate (right).

The quantum full adder is not superior to a classical full adder, but it does illustrate the difference between quantum and classical gates. Classical gates are physical objects, and classical bits are voltages that are passed from one gate to the next. In quantum circuits, the qubits are physical objects and quantum gates correspond to some manipulation of a qubit or multiple qubits. Depending on the type of qubit,

quantum gates might take the form of microwave pulses, laser pulses, or magnetic fields. One issue with quantum gates is gate fidelity, i.e. the percentage of the time that the gate performs the desired function. Because of noise and experimental uncertainty, the application of a particular gate process may not always produce the desired outcome.

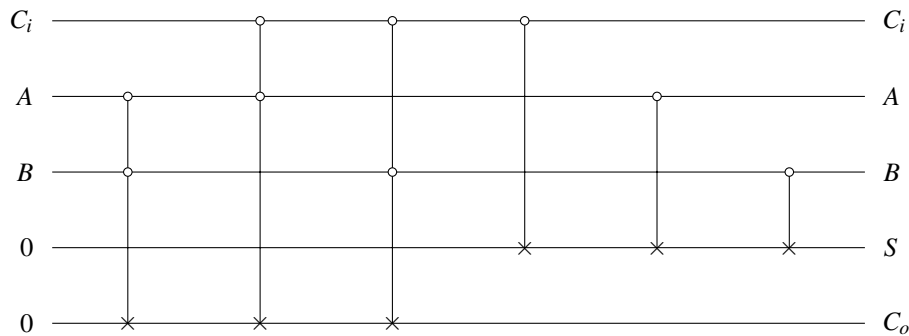


Figure 8.18 Quantum full adder circuit.

Delving further into quantum gates and their application in quantum algorithms is difficult without the use of linear algebra. For this reason, we will leave our discussion of quantum gates here, but there are a number of introductory textbooks on quantum computing for the reader who is interested in learning more.

8.8 PROBLEMS

- Problem 1.** (a) How many different possible values can an 8-bit binary number have?
(b) What is the decimal equivalent of the binary number 111001010?
(c) What is the binary equivalent of the decimal number 237?

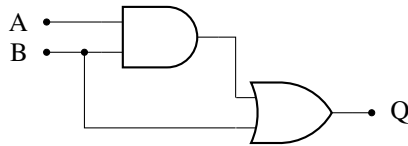
Problem 2. We digitize an analog signal that has a total duration of 7 minutes using a sampling rate of 30 kHz and with each amplitude value represented as a 6-bit number. How many megabytes (MB) is the resulting digital file?

Problem 3. An anti-aliasing filter is a low-pass filter used at the input of an analog-to-digital (A/D) converter that blocks frequencies that are beyond the maximum frequency that the A/D converter can accurately digitize. Briefly explain why this type of filter is useful. (It may be helpful to consider what would happen if, for example, you have a signal at frequency f_0 that is sampled at a rate of $f_0/3$.)

Problem 4. Briefly explain why coupling the output of a TTL digital circuit to the input of a 5V CMOS digital circuit could create problems, but coupling the output

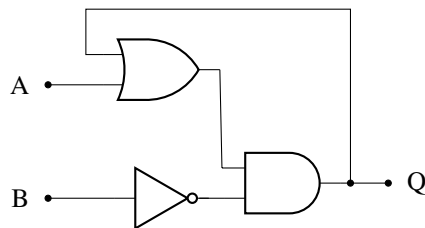
of a 5V CMOS digital circuit to the input of a TTL digital circuit should not create problems.

Problem 5. Determine the truth table for the circuit shown.

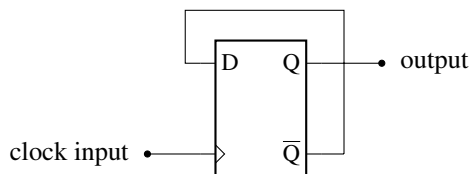


Problem 6. Design a circuit consisting of three AND gates that has four inputs A , B , C , and D and one output Q such that $Q = 1$ if all four inputs are 1 and $Q = 0$ otherwise.

Problem 7. Determine the truth table for the circuit shown.



Problem 8. Consider a D flip-flop in which the inverting output $\bar{Q}$ is fed back into the D input, as shown. Sketch both the clock input and the output Q as functions of time with equivalent time axes showing four clock periods. Assume that the circuit is edge-triggered, i.e. the output updates on each rising and falling edge of the clock. (Note that crossing wires are only connected if indicated by a black dot at the point of crossing.)



A Some Useful Constants, Conversions, Identities, Metric Prefixes, and Units

Table A.1

Some useful constants expressed in SI units with four significant figures

Constant	Symbol	Value
fundamental charge	q_0	$1.602 \times 10^{-19} \text{ C}$
electron mass	m_e	$9.109 \times 10^{-31} \text{ kg}$
Planck constant	h	$6.626 \times 10^{-34} \text{ J}\cdot\text{s}$
reduced Planck constant	$\hbar = h/(2\pi)$	$1.055 \times 10^{-34} \text{ J}\cdot\text{s}$
Boltzmann constant	k_B	$1.381 \times 10^{-23} \text{ J/K}$
permittivity of free space	ϵ_0	$8.854 \times 10^{-12} \text{ F/m}$
permeability of free space	μ_0	$1.257 \times 10^{-6} \text{ H/m}$
speed of light in free space	$c = 1/\sqrt{\epsilon_0\mu_0}$	$2.998 \times 10^8 \text{ m/s}$
impedance of free space	$Z_{fs} = \sqrt{\mu_0/\epsilon_0}$	$376.7 \text{ }\Omega$

Table A.2

Common metric prefixes

Prefix	Abbreviation	Value
femto-	f-	10^{-15}
pico-	p-	10^{-12}
nano-	n-	10^{-9}
micro-	μ -	10^{-6}
milli-	m-	10^{-3}
centi-	c-	10^{-2}
kilo-	k-	10^3
mega-	M-	10^6
giga-	G-	10^9
tera-	T-	10^{12}
peta-	P-	10^{15}

Table A.3
Some useful conversions and identities

1 eV = 1.602 × 10⁻¹⁹ J

1 gauss = 10⁻⁴ tesla

cos A cos B = $\frac{1}{2} \cos (A - B) + \frac{1}{2} \cos (A + B)$

sin A sin B = $\frac{1}{2} \cos (A - B) - \frac{1}{2} \cos (A + B)$

cos (tan⁻¹ x) = 1/√(1 + x²)

e^x = $\sum_{n=0}^{\infty} \frac{x^n}{n!}$

e^{ix} = cos(x) + i sin(x)

Table A.4
SI units for common quantities

Quantity	Symbol	Common SI Unit(s)	Fundamental SI Unit(s)
capacitance	<i>C</i>	farad (F)	m ⁻² ·kg ⁻¹ ·s ⁴ ·A ²
charge	<i>Q</i> ^a	coulomb (C)	s·A
current	<i>I</i>	ampere (A)	A
current density	$\vec{J}$	ampere per square meter (A/m ²)	m ⁻² ·A
electric field	$\vec{E}$	volt per meter (V/m)	m·kg·s ⁻³ ·A ⁻¹
energy	<i>E</i>	joule (J)	m ² ·kg·s ⁻²
frequency (angular)	ω	radian per second (rad/s) ^b	s ⁻¹
frequency (linear)	<i>f</i>	hertz (Hz)	s ⁻¹
inductance	<i>L</i>	henry (H)	m ² ·kg·s ⁻² ·A ⁻²
impedance	<i>Z</i>	ohm (Ω)	m ² ·kg·s ⁻³ ·A ⁻²
magnetic field	<i>B</i>	tesla (T)	kg·s ⁻² ·A ⁻¹
mass	<i>m</i>	kilogram (kg)	kg
power	<i>P</i>	watt (W)	m ² ·kg·s ⁻³
resistance	<i>R</i>	ohm (Ω)	m ² ·kg·s ⁻³ ·A ⁻²
resistivity	ρ	ohm meter (Ω·m)	m ³ ·kg·s ⁻³ ·A ⁻²
speed, velocity	<i>v</i> , $\vec{v}$	meter per second (m/s)	m·s ⁻¹
temperature	<i>T</i>	kelvin (K)	K
time	<i>t</i>	second (s)	s
voltage	<i>V</i>	volt (V)	m ² ·kg·s ⁻³ ·A ⁻¹

^a The symbol *Q* is also used for the quality factor of a resonator.
^b Note that radians are dimensionless.

B Review of Complex Numbers

The imaginary number is the square root of negative one. In the fields of math and science, the imaginary number is usually written as i , while in engineering it is often written as j . Here we will use i . The imaginary number squared is equal to negative one.

A complex number is a number that can be represented as the sum of two parts, one of which is purely real (i.e. it contains no i terms) and one of which is purely imaginary (i.e. it can be expressed as some purely real number multiplied by i). We can always write a complex number Z in the form $Z = X + iY$, where the real part X and the imaginary part Y are both themselves purely real numbers. For example, if $Z = 7 - i5$, then the real part $X = 7$ and the imaginary part $Y = -5$.

To add two complex numbers, we add their real parts together and add their imaginary parts together. For example, $(3 - i2) + (5 + i7) = (3 + 5) + i(-2 + 7) = 8 + i5$. Subtraction follows the same approach.

To multiply two complex numbers, we foil out the expression, making use of the fact that $i^2 = -1$. For example, $(3 - i2)(5 + i7) = 3 \times 5 + 3 \times i7 - i2 \times 5 - i2 \times i7 = 15 + i21 - i10 + 14 = 29 + i11$.

When we divide two complex numbers, we get an expression with a complex denominator. The complex conjugate is a useful concept for converting the denominator into a purely real number. The complex conjugate of a complex number is equal to the original complex number but with each i in the original number replaced by $-i$. For example, the complex number $Z = 4 + i3$ has a complex conjugate $Z^* = 4 - i3$. The product of a complex number Z and its complex conjugate Z^* is written as $|Z|^2$, which is called the “complex conjugate squared.” The complex conjugate squared is always a real number.

If we have a fraction with a complex denominator, we can make the denominator purely real by multiplying both the numerator and denominator by the complex conjugate of the denominator. This allows us to then separate the fraction into its real and imaginary parts. Consider the following example in which we assume A , B , C , and D are real numbers:

$$\frac{A + iB}{C + iD} = \left(\frac{A + iB}{C + iD} \right) \left(\frac{C - iD}{C - iD} \right) = \frac{AC + BD - iAD + iBC}{C^2 + D^2} = \frac{AC + BD}{C^2 + D^2} + i \frac{BC - AD}{B^2 + C^2}$$

We can represent our complex number $Z = X + iY$ as a two-dimensional vector on the real-imaginary plane, in analogy to a two-dimensional vector on the $\mathbf{x}$ - $\mathbf{y}$ plane. An example of this for $X = -3$ and $Y = 2$ is shown in [Figure B.1](#). The length, or magnitude, of this vector is $|Z| = \sqrt{X^2 + Y^2}$ and the angle ϕ relative to the real axis is $\phi = \tan^{-1}(Y/X)$. Because of the tangent function, ϕ is constrained to $-\pi/2 \leq \phi \leq$

$\pi/2$, and so ϕ is measured from either the positive or negative real axis, whichever is closest, with a positive angle corresponding to a counterclockwise displacement from the real axis and a negative angle corresponding to a clockwise displacement.

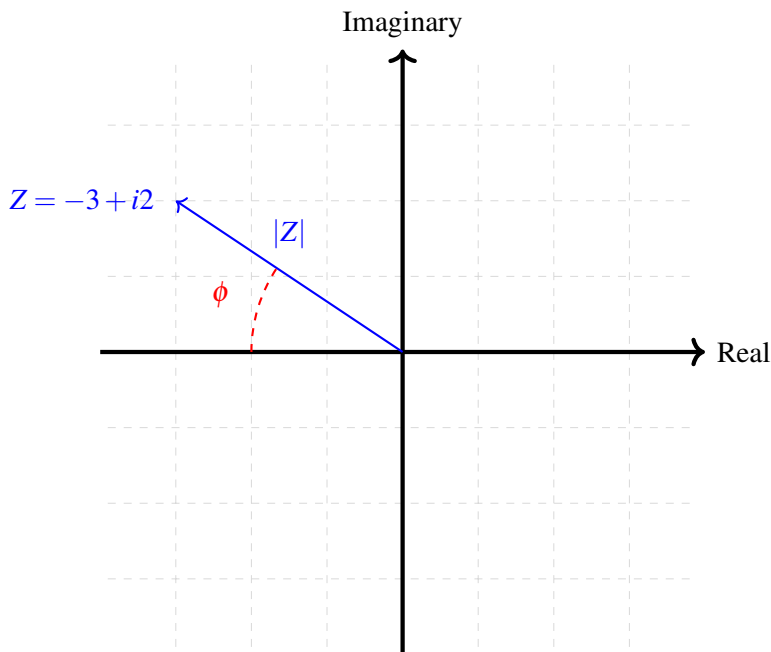


Figure B.1 Two-dimensional vector representation of a complex number $Z = -3 + i2$ which has magnitude (length) $|Z| = \sqrt{3^2 + 2^2} = 3.61$ and angle $\phi = \tan^{-1}(-2/3) = -33.7^\circ$.

A complex number can also be written in the form $Z = |Z|e^{i\phi}$, which is known as an exponential or polar representation. In this form, it is straightforward to multiply and divide complex numbers. For example, $(Ae^{i\phi_1})(Be^{i\phi_2}) = AB e^{i(\phi_1 + \phi_2)}$ and $(Ae^{i\phi_1}) / (Be^{i\phi_2}) = (A/B) e^{i(\phi_1 - \phi_2)}$. To convert to the form $Z = X + iY$, one can make use of Euler's formula, $e^{i\phi} = \cos \phi + i \sin \phi$.

Index

- active region, 88
- adder, binary, 129-130
- adder, op-amp, 103-104
- ammeter, 13-14
- amplifier noise, 119-120
- amplitude, root-mean-square (rms), 34
- amplitude, zero-to-peak, 34
- analog-to-digital conversion, 126-127
- anharmonic oscillator, 141
- antenna *see* [dipole antenna](#)
- attenuation constant, 62

- balun, 69-70
- bandgap, 71-74
- battery, 4
- binary, 125
- binary counter, 137-138
- bipolar junction transistor, 86-89
- bit, 125
- blackbody radiation, 118
- Bloch sphere, 141-142
- BNC connector, 39-40
- Bode plot, 41-42
- Boltzmann constant, 72
- byte, 125

- capacitance, 23-24
- capacitor *see* [capacitance](#)
- characteristic impedance, 63
- chemical potential, 71-73
- clock, digital, 134
- coaxial cable, 59, 63
- common emitter amplifier, 93-95
- common emitter configuration, 87-88
- comparator, 106-108
- conductance, 1
- conductivity, 2-3
- coupling capacitor, 93-94
- current, 1
- current density, 2-3
- current source, 4, 13, 32

- D flip-flop, 135-136
- Darlington pair, 91-93
- depletion region, 77-78, 87-88
- dielectric constant, 23
- differential amplifier, 94-95
- digital filtering, 121-123
- digital multimeter (DMM) *see* [multimeter](#)
- diode *see* [PN junction](#)
- dipole antenna, 67-70
- doping, 73-75
- drift velocity, 2-3
- Drude model, 2-3
- duty cycle, 134

- electric field, 1-3
- electric potential difference *see* [voltage](#)
- emitter follower, 90-91
- entangled states, 139-140, 143-144

- far field, 69
- feedback, stable, 102
- feedback, unstable, 106-107
- Fermi energy, 71-72
- Fermi-Dirac distribution, 72
- ferrite core, 55
- field-effect transistor, 86-87
- filters, 45-49
- filters, active, 105
- flicker noise *see* [one-over-f noise](#)
- forward bias, 76-79
- Fourier series, 50-51
- Fourier transform, 51-54
- four-wire measurement, 14
- frequency mixing, 84-86
- full-wave rectifier, 82-84
- fundamental charge, 1

- ground, 8

- half-wave rectifier, 82
- harmonic oscillator, 141

- heterodyne mixer *see* [frequency mixing](#)
- hole, [73](#)
- homodyne mixer *see* [frequency mixing](#)

- ideal op-amp assumptions, [102-103](#)
- impedance, [30-33](#)
- inductance, [24-25](#), [32](#)
- inductor *see* [inductance](#)
- input bias current, [106](#)
- input impedance, [65-66](#)
- input offset current, [106](#)
- input offset voltage, [105-106](#)
- integrated circuit, [95-97](#)
- inverting amplifier, [103-104](#)

- JK flip-flop, [136-137](#)
- Johnson noise thermometry, [117](#)
- Johnson-Nyquist noise, [116-118](#)

- Kirchhoff's rules, [10-12](#)

- LC circuit, [28-30](#)
- lock-in amplifier, [86](#)
- logic gates, [127-129](#)
- logic, 5V CMOS, [125-126](#)
- logic, TTL, [125-126](#)
- lumped element, [60](#)

- mixer *see* [frequency mixing](#)
- multimeter, [14](#)

- near field, [69](#)
- noise current spectral density, [116](#)
- noise temperature, [120](#)
- noise voltage spectral density, [115-116](#)
- non-inverting amplifier, [103](#)
- Norton equivalent circuit, [16-17](#)

- ohmmeter, [14](#)
- Ohm's law, [1](#)
- one-over-f noise, [121](#)
- open-loop gain, [101](#)
- oscilloscope, [38-41](#)

- parallel capacitors, [24](#)
- parallel impedances, [32](#)

- parallel inductors, [25](#)
- parallel resistors, [6](#)
- phase angle, [31](#)
- phase constant, [62](#)
- phase velocity, [62](#)
- pickup noise, [121](#)
- Planck constant, [72](#)
- PN junction, [75-79](#)
- potentiometer, [8](#)
- power in AC circuits, [33-34](#)
- power in DC circuits, [9](#)

- quality factor, [35-38](#)
- quantum circuits, [144-145](#)
- quantum gates, [141-144](#)
- quarter-wave transformer, [66](#)
- qubits, [138-141](#)

- radiation pattern, [69](#)
- RC circuit, [26-27](#)
- reflection coefficient, [64-65](#)
- relative permeability, [25](#)
- relative permittivity *see* [dielectric constant](#)
- relaxation oscillator, [110-111](#)
- resistance, [1-3](#), [5-6](#)
- resistivity, [2](#)
- resistor *see* [resistance](#)
- resonant frequency, [29-30](#), [35-37](#)
- reverse bias, [76-79](#)
- ripple, [121](#)
- RL circuit, [28](#)
- RLC circuit, [30](#), [34-38](#)

- saturation, [94](#), [101-102](#)
- saturation region, [88](#)
- scattering parameters, [66-67](#)
- Schmitt trigger, [107-110](#)
- series capacitors, [24](#)
- series impedances, [32](#)
- series inductors, [25](#)
- series resistors, [5](#)
- set-reset latch, [130-134](#)
- shift register, [136](#)
- Shot noise, [118-119](#)

- signal-to-noise ratio, 115
- slew rate, 105-106
- switch bounce, 132-134
- switching noise, 106-108
- terminal, 4
- Thevenin equivalent circuit, 16-18
- threshold voltage, 77, 80
- time constant, 27-28
- transfer function, 45
- transformer, 54-55
- transistor current source, 89-90
- voltage, 1
- voltage clamp, 80-81
- voltage divider, 6, 45-46
- voltage source, 4, 12-13, 32
- voltage standing wave ratio (VSWR), 70
- voltmeter, 13, 15
- white noise, 116
- Y-factor measurement, 120
- Zener diode, 79